The Network Horizon

A General Theory of Polycentric Technological Emergence

Network Horizon Series, Volume V

Silvano Salvador

Contents

1 The Problem of Convergence **6**

1.1 A Recurrent Pattern in Prehistory 6
1.2 Why Diffusionism Became the Default 7
1.3 The Epistemology of Independent Invention 8
1.4 From Empirical Findings to Theoretical Questions 8
1.5 The Vocabulary of the Series: A Retrospective Clarification 9
1.6 Plan of the Volume 10
1.7 The Persistence of Diffusionism and Its Institutional Roots 11
1.8 The Comparative Scope of the Series: What It Covers and What It Does Not . 12
1.9 The Asymmetry of Burden of Proof Reconsidered 13
1.10 Calibration Against Known Cases of Both Types 15
1.11 The Demarcation Problem: What Counts as a Technology for the Purposes of This Series 16
1.12 The Status of Negative Cases: Technologies That Fail the Evidential Completeness Criterion 17
1.13 Null Models and the Detection of Chance Convergence 18
1.14 The Role of Analogical Reasoning from Ethnographic and Historical Cases . . 21

2 Axiomatic Foundations of the Network Horizon Model **23**

2.1 The Need for Explicit Axiomatisation 23
2.2 Primitive Concepts 23
2.3 The Axioms 24
2.4 Derived Structural Conditions 25
2.5 The Parameter Space of the Model 26
2.6 Relationship to Existing Theoretical Frameworks 28
2.7 A Note on the Wheel as a Limiting Case 30
2.8 Summary of the Axiomatic Structure 30
2.9 The Ontological Status of Axioms in Archaeological Theory 31
2.10 The Parameter ρ^* and Its Estimation 32
2.11 Extensions of the Axiomatic Framework 33
2.12 Worked Derivation: Computing τ_{inv} from Axiom-Level Parameters 34
2.13 The Network Horizon in Formal Analogy to Percolation Theory 35
2.14 Counterfactual Robustness of the Axiomatic Structure 37
2.15 The Relationship Between Axiom Satisfaction and Effect Size 38
2.16 Extending the Axiom Set: A Cautionary Note on Proliferation 39
2.17 The Logical Form of the Axiom Set: Conjunction Versus Weighted Combination 39
2.18 Formal Properties of the Composite Score Under the Product Rule 40
2.19 An Extended Worked Derivation: From Raw Evidence to Composite Score . . 41
2.20 The Treatment of Missing Evidence Within the Axiomatic Scoring Protocol . . 42
2.21 Correlated Axiom Violations and the Limits of the Product Rule 43

2.22 The Role of Cognitive Science in Grounding Axiom II 46

3 Cross-Case Synthesis: The Four Technologies as Instances of a Single Structural Pattern 48
3.1 The Logic of Retrospective Axiomatic Application 48
3.2 Ceramics (Volume IV) 49
3.3 Horse Domestication (Volume I) 50
3.4 Copper Metallurgy (Volume III) 51
3.5 The Wheel and Wheeled Transport (Volume II) 53
3.6 Comparative Synthesis 54
3.7 The Ceramics Case as the Framework's Most Stringent Test 55
3.8 The Temporal Structure of Evidence Accumulation Across the Four Cases . . 56
3.9 Quantifying Inter-Technology Structural Similarity 57
3.10 The Wheel Revisited: A Quantitative Re-evaluation 58
3.11 The Wheel as a Methodological Stress-Test: Implications for Monocentric Critiques 59
3.12 The Ceramics Case as a Template for Future Multi-Source Analysis 62
3.13 Non-Additive Interactions Among Axioms 62
3.14 Comparative Timescales of Independent Invention Events 63
3.15 The Copper Metallurgy Case: Stratigraphic Decomposition of the Evidence . 64
3.16 The Horse Domestication Case: The Botai Debate and Its Axiomatic Resolution 65
3.17 The Ceramics Case: Regional Traditions in Depth 66
3.18 Axiomatic Assessment and the Role of Expert Disagreement 67
3.19 Cross-Case Comparison of Evidentiary Stream Contribution 67
3.20 The Problem of Evidence Stream Correlation Under Shared Sampling Constraints 69
3.21 Quantitative Comparison with Alternative Polycentric Frameworks 69
3.22 A Formal Measure of Cross-Case Structural Similarity 70
3.23 Structural Similarity and the Question of Shared Underlying Causation 71
3.24 Publication Bias and the Selection History of the Four-Case Corpus 72
3.25 A Comparative Timeline Visualisation Across the Four Cases 74

4 The Diffusion Boundary Problem 77
4.1 Locating ρ^* Empirically 77
4.2 Diagnostic Test Suite 77
4.2.1 Chronological Overlap and the Minimum Diffusion Transit Time . . . 78
4.2.2 Material Provenance and the Zero-Contact Diagnostic 78
4.2.3 Population Genetic Evidence and Admixture-Free Attestation 79
4.2.4 Linguistic Vocabulary Independence and Diffusion Borrowing Signatures 80
4.3 Combining the Diagnostics: A Bayesian Integration Framework 80
4.4 The Diffusion Boundary in Continuous Space 82
4.5 Implications for Future Empirical Work 83
4.6 The Contact-Network as a Probabilistic Graph 84
4.7 Sensitivity of the Posterior to the Choice of Prior 85
4.8 The Diffusion Velocity Function and Its Archaeological Calibration 86
4.9 Mixed-Regime Evidence and Its Interpretation 87
4.10 The Empirical Evidence for Diffusion Transit Rates 88
4.11 When Identical Evidence Supports Different Conclusions 89
4.12 Network Topology Beyond Density: The Role of Structural Position 90
4.13 Diffusion Under Heterogeneous Adoption Thresholds 91
4.14 Comparative Calibration: Documented Historical Diffusion Velocities 92
4.15 Seasonal and Logistical Constraints on Pedestrian Diffusion Velocity 92

4.16 The Interaction Between Network Density and Technology-Specific Tacitness Revisited . . . 94
4.17 Maritime and Riverine Corridors: A Differentiated Transmission Model . . . 95
4.18 Sensitivity of the Network Horizon to Population Growth Dynamics . . . 97

5 Methodological Apparatus of the Network Horizon Framework 99
5.1 Purpose of a Unified Methodological Critique . . . 99
5.2 Bayesian Sequence Modelling of Radiocarbon Dates . . . 99
5.3 Instrumental Neutron Activation Analysis (INAA) and Material Provenance 100
5.4 Ancient DNA and Population Genetic Evidence . . . 101
5.5 Computational Phylogenetic Glottochronology . . . 102
5.6 The Architecture of Convergent Evidence . . . 103
5.7 What the Methodological Apparatus Cannot Do . . . 104
5.8 The Problem of Method Interdependence in Multi-Source Analysis . . . 104
5.9 The Role of Negative Evidence . . . 105
5.10 Detection of Subtle Contact: The Sub-Threshold Problem . . . 106
5.11 The Chronometric Resolution Problem for Future Cases . . . 107
5.12 The *Chaîne Opératoire* as a Falsifying Tool . . . 107
5.13 Integrating Ethnoarchaeological Evidence . . . 109
5.14 Inter-Rater Reliability in Axiom Scoring . . . 109
5.15 Sensitivity of Conclusions to Expert Disagreement Resolution Procedures . . . 111
5.16 The Role of Disconfirming Expert Opinion . . . 111
5.17 Negative Evidence Revisited: Formal Treatment of Absence . . . 112
5.18 Computational Reproducibility and the Audit Trail . . . 112
5.19 Comparative Evaluation of Alternative Statistical Combination Frameworks . . . 114

6 Falsifiability and Predictive Scope 117
6.1 A Theory Without Falsifiable Predictions Is Not a Theory . . . 117
6.2 Structural Predictions: Technologies Expected to Show Polycentric Emergence 118
6.3 Diagnostic Predictions: What Future Evidence Should Find . . . 119
6.4 Conditions for Falsification of the Model as a Whole . . . 121
6.5 Positioning the Network Horizon Model in the Theoretical Landscape . . . 122
6.6 The Problem of Confirmation Bias in Prospective Prediction . . . 123
6.7 The Iron Smelting Case in More Detail . . . 124
6.8 The Predictive Scope Beyond Technology . . . 126
6.9 Failure Modes of the Prospective Prediction System . . . 126
6.10 The Relationship Between Predictive Success and Theoretical Depth . . . 127
6.11 A Worked Prospective Case: Early Pottery Firing Technique Variation . . . 128
6.12 Calibrating Prediction Confidence Against Track Record . . . 129
6.13 Distinguishing Predictive Success from Retrodictive Fit . . . 129
6.14 A Second Worked Prospective Case: Textile Loom Technology . . . 130
6.15 Comparing the Two Worked Prospective Cases: A Methodological Reflection 131
6.16 A Third Worked Prospective Case: Basketry and Plant-Fibre Technology . . . 132
6.17 A Fourth Falsification Class: Predictions About Future Archaeological Discoveries . . . 134

7 Limits, Open Questions, and the Horizon Ahead 137
7.1 What This Series Has and Has Not Established . . . 137
7.2 The Inference from Four Cases to a General Pattern . . . 139
7.3 The Three Most Productive Open Questions . . . 139

7.4 A Note on the Significance of Polycentric Emergence for Understanding Human Cognition . . . 140
7.5 Closing Remarks . . . 141
7.6 The Boundary Between This Series and Its Successors . . . 142
7.7 The Epistemological Status of the Network Horizon's Central Claim . . . 143
7.8 The Relationship Between Polycentric Emergence and Human Behavioural Modernity . . . 143
7.9 A Critical Engagement with the Series' Own Methodology . . . 144
7.10 The Network Horizon as a Model for Multi-Disciplinary Integration . . . 145
7.11 Final Accounting: What Remains Unknown . . . 146
7.12 Implications for the Philosophy of Archaeological Inference . . . 147
7.13 The Series in Relation to Falsificationist and Bayesian Philosophy of Science . . . 148
7.14 Anticipated Objections Not Yet Addressed . . . 148
7.15 Closing Reflection on the Series as a Whole . . . 149
7.16 A Final Note on the Relationship Between Formal Rigour and Archaeological Interpretation . . . 149
7.17 Toward a General Theory of Technological Convergence: Speculative Extensions Beyond Archaeology . . . 150
7.18 A Final Methodological Postscript: On Writing a Synthesis Volume After Four Case-Specific Volumes . . . 152

A Consolidated Axiomatic Profiles of the Four Core Cases **155**
A.1 Qualitative Satisfaction Ratings by Axiom . . . 155
A.2 Composite Scores Cited in This Volume . . . 156

B A Candidate Sixth Axiom: Institutional Prerequisite **157**
B.1 Motivation: A Gap the Five Axioms Do Not Cover . . . 157
B.2 Why Iron Smelting Suggests the Need for the Distinction . . . 158
B.3 Why This Is Offered as an Appendix, Not a Sixth Axiom in Chapter 2 . . . 158

References **160**

Chapter 1

The Problem of Convergence

1.1 A Recurrent Pattern in Prehistory

The history of technology as conventionally narrated is a history of origins: single points of invention, followed by diffusion outward along trade routes, migration corridors, or lines of cultural contact. This narrative has the virtue of parsimony in its simplest form — one cause, many effects — and it has structured the dominant interpretive tradition in archaeology, historical linguistics, and the history of science for more than a century. It is also, as the four preceding volumes of this series have argued in detail, systematically wrong for the class of technologies that shaped the Neolithic and Chalcolithic world.

The empirical record assembled in Volumes I through IV documents a recurrent pattern: technologies of fundamental adaptive importance — ceramic vessels, the wheel, horse domestication, copper metallurgy — appear in multiple geographically and culturally separated contexts within chronological windows that are too narrow, and across distances too great, to be explained by any plausible diffusion mechanism. The pattern is not occasional or marginal. It is the dominant signature of the prehistoric technological record once the evidence is subjected to rigorous, multi-source analysis rather than evaluated through the prior of single-origin diffusionism.

This volume does not add a fifth empirical case to the four already established. Its purpose is different and, in a sense, more ambitious: to ask what the recurrence of this pattern implies about the structure of human technological cognition and social organisation, to formalise the conditions under which polycentric emergence is structurally predictable rather than merely possible, and to state explicitly what the Network Horizon model is — its axioms, its scope, its limits, and its falsifiable predictions — so that it can be evaluated, criticised, and extended by future research rather than remaining an implicit framework distributed across four separate monographs.

Definition 1.1: The Polycentric Emergence Problem

Let $\mathcal{T}$ be a technology defined by a minimally sufficient set of functional and material criteria $\{c_1, \dots, c_k\}$ (the *technological specification*). Let $\{R_1, \dots, R_n\}$ be a set of $n \geq 2$ geographic regions, each associated with an earliest secure attestation date τ_i for $\mathcal{T}$ and a demographic history $\mathcal{H}_i$. The *polycentric emergence problem* for $\mathcal{T}$ is to determine whether the observed distribution $(\tau_1, \dots, \tau_n; \mathcal{H}_1, \dots, \mathcal{H}_n)$ is better explained by:

(D) a *diffusion model*: a single invention event at time $\tau^* = \min_i \tau_i$ in region $R^* = \arg\min_i \tau_i$, followed by transmission to the remaining $n - 1$ regions at rates consistent with known contact-network parameters; or

(P) a *polycentric model*: $m \geq 2$ independent invention events, each occurring in response

to locally recurrent conditions $\mathcal{E}_i \subseteq \mathcal{H}_i$, without requiring inter-regional transmission of the technology prior to local invention.

The problem is non-trivial because both models can in principle generate any observed distribution $(\tau_1, \ldots, \tau_n)$; discrimination requires auxiliary evidence from material culture, archaeometry, population genetics, and historical linguistics, each contributing to the posterior probability of (D) versus (P) under a Bayesian evidence-integration framework.

1.2 Why Diffusionism Became the Default

The dominance of diffusionist interpretations in prehistoric archaeology is not an accident of intellectual history; it reflects a genuine prior probability argument. If we know nothing about the relative frequency with which major technologies are independently invented versus diffused, the assumption that complex technologies are invented rarely and spread widely is conservative in the sense of requiring fewer independent causal events. Invoking independent invention twice requires two sufficient-cause sequences; invoking diffusion once requires only one. Under a principle of parsimony applied naively to the count of causal events, diffusion is always the simpler hypothesis.

The error in this argument is not logical but empirical: it miscounts the causal events required by each model. A diffusion explanation for a technology appearing in two widely separated regions at similar dates does not require one causal event but a chain of them — the original invention, the transmission mechanism, the reception context, the adaptation of the technology to local materials and practices, and crucially, the non-occurrence of independent local invention during the transmission interval. When the transmission distance is large, the chronological gap between attestations is small, and the local ecological and material conditions are sufficient for independent invention, the diffusion model requires a more complex causal chain, not a simpler one. The polycentric model requires fewer auxiliary assumptions precisely in those cases where local conditions make independent invention structurally expectable.

Proposition 1.1: The Parsimony Inversion

Let C_D and C_P denote the minimum number of independently specified causal conditions required to generate the observed evidence **E** under models (D) and (P) respectively. The standard argument for diffusionism assumes $C_D < C_P$ because (D) posits one invention event to (P)'s $m \geq 2$. This argument is valid only when the following conditions hold simultaneously:

(i) a plausible contact mechanism exists between all attested regions within the observed chronological window;

(ii) the material, archaeometric, and genetic evidence is consistent with technology transfer along that mechanism;

(iii) local conditions in the secondary regions are insufficient for independent invention without external input.

When conditions (i)–(iii) fail jointly — as they do for each of the four cases documented in Volumes I–IV — the inequality reverses: $C_P < C_D$, and the polycentric model is the parsimonious one. The naive parsimony argument for diffusionism assumes, without empirical verification, that conditions (i)–(iii) hold; the empirical programme of this series consists precisely in testing those assumptions and finding them violated.

1.3 The Epistemology of Independent Invention

A persistent asymmetry in the archaeology of technology is that diffusion is usually the *default* hypothesis, while independent invention is treated as requiring affirmative proof. This asymmetry is epistemically unjustified, and its consequences have been substantial: plausible independent-invention hypotheses have been dismissed for lack of proof of independence, while diffusionist hypotheses have been accepted without proof of contact. The appropriate null hypothesis depends on the prior probability of the mechanisms involved, not on a convention that favours one model over the other.

Definition 1.2: Evidential Asymmetry and the Appropriate Null

In testing (D) versus (P) for a given technology $\mathcal{T}$ and region pair (R_i, R_j), the appropriate null hypothesis is determined by the prior probability ratio π_D/π_P, where:

$$\pi_D \propto \Pr(\text{contact}(R_i, R_j) \mid \mathcal{H}_i, \mathcal{H}_j, \tau_i, \tau_j) \cdot \Pr(\text{transmission} \mid \text{contact})$$

$$\pi_P \propto \Pr(\mathcal{E}_i \text{ sufficient} \mid \mathcal{H}_i) \cdot \Pr(\mathcal{E}_j \text{ sufficient} \mid \mathcal{H}_j)$$

When $\pi_P > \pi_D$ — that is, when local conditions in both regions independently exceed the sufficiency threshold for invention, and when contact probability is low — the null hypothesis is polycentric emergence, not diffusion. The burden of proof lies with the diffusion model, which must adduce positive evidence of contact and transmission rather than merely the absence of proof of independence.

The implications of this reframing are significant. It is not enough to note that diffusion is *possible* in order to prefer it over independent invention; diffusion must be shown to be *more probable than local invention* given the full evidential record. This is a substantially higher bar than the field has historically applied, and it is the bar that the Network Horizon series has consistently used.

1.4 From Empirical Findings to Theoretical Questions

The four volumes of this series have established, to a high degree of evidential confidence, that ceramic technology, wheeled transport, horse domestication, and copper smelting each satisfy the polycentric emergence criterion of Definition 1.1 for at least two and in most cases more than two independent regional traditions. The empirical programme is, in this sense, complete.

What it does not yet provide is an answer to the second-order question: *why* do these particular technologies show this particular pattern? Is polycentric emergence a general feature of all major prehistoric technologies, or is it specific to the technologies examined here? If specific, what properties of those technologies make them structurally prone to independent invention? And if general, what does this imply about models of human cognitive and social evolution that have treated diffusion as the primary engine of prehistoric technological change?

Definition 1.3: Structural Proneness to Polycentric Emergence

A technology $\mathcal{T}$ is *structurally prone* to polycentric emergence if the conditions $\mathcal{E}$ sufficient for its independent invention are satisfied in multiple geographically separated regions within a chronological window $\Delta\tau$ shorter than the expected diffusion transit

time τ_{diff} between those regions, i.e.:

$$\Pr(\mathcal{E}_i \text{ sufficient}, \mathcal{E}_j \text{ sufficient}, \Delta\tau < \tau_{\text{diff}}) > \epsilon$$

for some non-negligible threshold $\epsilon > 0$. Structural proneness is a property of the technology and its ecological-material context, not of any specific historical realisation. A technology that is structurally prone will *tend* to show polycentric emergence across the full range of prehistoric contexts in which its sufficient conditions are met, not merely in the specific cases documented in the archaeological record.

The central claim of this volume is that the four technologies examined in this series share a set of structural properties — formalised in Chapter 2 as the axioms of the Network Horizon model — that jointly constitute structural proneness. These properties are not unique to the four cases; they are shared, in varying degrees, by a wider class of technologies whose prehistoric record has not yet been subject to the same quality of multi-source analysis. The Network Horizon model therefore has predictive scope beyond the cases that motivated it, a claim that is examined critically in Chapter 6.

1.5 The Vocabulary of the Series: A Retrospective Clarification

The term "Network Horizon" has been used across the four preceding volumes in two related but distinct senses that it is now useful to distinguish explicitly. In its narrow, regional sense, a *network horizon* is an archaeologically defined threshold above which material evidence of inter-site interaction is detectable — the point at which a regional network of sites becomes sufficiently integrated to transmit technologies, styles, or people. In its broader, theoretical sense — the sense that gives this series its name — the Network Horizon is the boundary between the regime in which diffusion is the dominant mode of technological spread and the regime in which local invention recurs faster than diffusion propagates. Technologies that emerge before or below the Network Horizon — before regional interaction networks become dense enough to carry them — show polycentric signatures; technologies that emerge after or above it show diffusion signatures.

Definition 1.4: The Network Horizon (Formal Statement)

Let $\rho(t)$ denote the density of inter-regional contact links in a population network at time t, and let $\tau_{\text{inv}}(\mathcal{T})$ denote the expected time between independent inventions of technology $\mathcal{T}$ given local sufficient conditions. The *Network Horizon* for $\mathcal{T}$ is the critical density $\rho^*(\mathcal{T})$ such that:

$$\rho^*(\mathcal{T}) = \inf\{\rho : \tau_{\text{diff}}(\mathcal{T}, \rho) < \tau_{\text{inv}}(\mathcal{T})\}$$

that is, the minimum contact density above which diffusion of $\mathcal{T}$ outpaces its independent re-invention. For $\rho < \rho^*(\mathcal{T})$, the expected number of independent invention events exceeds one, and polycentric emergence is the structurally predicted outcome. For $\rho > \rho^*(\mathcal{T})$, diffusion from the first invention site reaches all other sites before independent re-invention occurs, and single-origin diffusion is the structurally predicted outcome.

This formalisation has an important corollary: the Network Horizon is technology-specific. A technology with a low τ_{inv} (easy to independently re-invent given the right conditions) has a lower ρ^* than a technology with a high τ_{inv} (difficult to re-invent). The same regional contact network may be above the Network Horizon for one technology and below it for another. This explains why polycentric emergence and single-origin diffusion can coexist in the same archaeological period and the same geographic region: different technologies have different

ρ^* values, and the same network density places them on different sides of their respective horizons.

Corollary 1.1: Coexistence of Diffusion and Polycentric Emergence in the Same Period

For a given period $[t_1, t_2]$ and region set $\{R_1, \ldots, R_n\}$ with contact density ρ, technologies $\mathcal{T}_a$ and $\mathcal{T}_b$ with $\rho^*(\mathcal{T}_a) > \rho > \rho^*(\mathcal{T}_b)$ will show polycentric emergence for $\mathcal{T}_a$ and single-origin diffusion for $\mathcal{T}_b$ simultaneously. This coexistence is not a contradiction but the expected consequence of technology-specific variation in ρ^*. Archaeological periods that show mixed evidence — some technologies diffusing, others independently invented — are not anomalies requiring special explanation; they are the predicted output of a world in which different technologies sit on different sides of the same contact-density threshold.

1.6 Plan of the Volume

Chapter 2 formalises the axioms of the Network Horizon model, deriving from them the structural conditions that make a technology prone to polycentric emergence. Chapter 3 applies those axioms retrospectively to the four cases of Volumes I–IV, demonstrating that all four satisfy the axiomatic conditions and identifying where each case sits within the model's parameter space. Chapter 4 develops the Diffusion Boundary Problem — the formal question of how to locate ρ^* empirically for a given technology — and proposes a suite of diagnostic tests. Chapter 5 provides a critical methodological synthesis: an evaluation of the evidentiary methods used across the series in terms of discriminating power, resolution limits, and optimal combination strategies. Chapter 6 derives the model's falsifiable predictions for technologies not yet examined and assesses the conditions under which future research could confirm or refute them. Chapter 7 acknowledges the limits of the framework and proposes the agenda for the research programme it opens rather than closes.

Note 1.0: What Kind of Claim This Volume Makes — Stated Before the Argument, Not After

Two qualifications govern everything that follows, and stating them here rather than leaving them to Chapter 7 is a deliberate choice: a reader should be able to calibrate the strength of every subsequent claim from the outset rather than reconstructing the appropriate level of confidence only after reaching the book's final pages.

First, the calibration corpus is four cases. Proposition 7.1 shows that with $n = 4$, the standard error on a predicted composite score is wide enough to flip a borderline case across the Regime I/II boundary, and that reducing this uncertainty to a level where fine-grained predictions are reliable requires $n \geq 10$–15 cases spanning the parameter space. Nothing in the axiomatic apparatus of Chapter 2 changes this; formalisation sharpens what the four cases say, it does not substitute for a fifth, sixth, or tenth case. Coarse classification (which of three well-separated bands a new technology falls into) is supported by the current corpus; precise point estimates are not, and the volume does not claim otherwise past Chapter 6.

Second, the model's central claim — that polycentric emergence is the structurally predicted outcome for technologies satisfying Axioms I–V — is what Definition 7.1 below calls a *structural prediction*, not an empirical finding in the sense that the four case-specific posterior odds of Chapter 3 are empirical findings. A structural prediction is supported or undermined gradually, by accumulation across many cases extending beyond this series; it is not confirmed by the four cases used to build it, and no single additional case, however striking, would confirm it either. Readers approaching this volume for

its theoretical architecture should hold the two registers apart: the evidentiary conclusions about ceramics, the wheel, horse domestication, and copper metallurgy carry the empirical weight established in Volumes I–IV; the general theory built on top of them in this volume carries the weaker, provisional status appropriate to a four-case research programme specification, and is offered as such.

1.7 The Persistence of Diffusionism and Its Institutional Roots

The foregoing epistemological analysis might suggest that diffusionism's dominance in archaeology is a simple matter of prior probabilities: researchers apply the most conservative hypothesis by default, update it when evidence demands, and otherwise maintain theoretical neutrality. The historical reality is more complex. Diffusionism did not merely emerge as a probabilistically justified prior; it became embedded in the institutional apparatus of archaeology during a formative period in which its proponents wielded substantial disciplinary authority, its methods were operationalised in research programmes and taught in university curricula, and its conclusions were repeated across enough secondary literature to acquire the appearance of settled fact. Understanding the persistence of diffusionism requires understanding not only its epistemic structure but its sociology.

The classic hyper-diffusionist positions of the late nineteenth and early twentieth centuries — Grafton Elliot Smith's Egyptian origins of civilisation, W.J. Perry's elaborations thereof, and the quasi-mythological narratives of cultural radiation from a single cradle — are now universally rejected, and rightly so. They failed not because diffusionism as a theoretical stance is wrong but because they were applied without methodological constraint, generating unfalsifiable genealogies of cultural resemblance that accumulated similarities regardless of whether those similarities were independent, transmitted, or merely superficial. The reaction against hyper-diffusionism in mid-twentieth-century archaeology was not, however, a move toward the principled weighing of diffusion against independent invention; it was, in many research traditions, a move toward a methodological agnosticism that declined to theorise the mechanisms of cultural spread at all. Processual archaeology in the 1960s and 1970s was concerned primarily with social and ecological processes internal to regional sequences, and while it deployed quantitative methods with greater rigour than its predecessors, it did not develop the multi-source evidentiary framework that would be needed to distinguish polycentric from diffusionist patterns in the record.

Note 1.1: The Processual Archaeology Interval and Its Legacy

The processual turn in archaeology (Binford 1962; Clarke 1968) is sometimes credited with introducing scientific method into a field previously dominated by culture-historical particularism. This is correct as far as it goes, but it obscures a significant limitation: processual archaeology's primary objects of explanation were adaptive trajectories within regional sequences, and its statistical methods (seriation, cluster analysis, multivariate ordination) were designed to track internal variation within tradition-linked assemblages rather than to test for inter-regional independence. The result was a research infrastructure well equipped to describe how technologies changed within a region but poorly equipped to ask whether similar technologies in different regions had a common origin. This infrastructure was not neutral with respect to the diffusion/polycentricity debate; it was structurally predisposed toward asking questions that presupposed the regional continuity of traditions, which in turn presupposed that those traditions had beginnings rather than that they arose independently from structurally similar local conditions. The

multi-source evidentiary framework developed across the Network Horizon series is, in part, a methodological response to the analytical gaps left by this processual legacy.

The post-processual turn of the 1980s and 1990s (Hodder 1982; Shanks and Tilley 1987) reintroduced questions of meaning, agency, and inter-regional connection into archaeological interpretation, and its emphasis on identity and material culture as vehicles of social signalling created intellectual space for renewed interest in long-distance exchange. But post-processual interest in connectivity was not, in general, an interest in the statistical discrimination between diffusion and independent invention; it was an interest in the symbolic and social dimensions of exchange networks that, by and large, presupposed the existence of those networks rather than testing whether specific technologies propagated along them. The default diffusionism of the archaeological mainstream thus survived the processual-to-post-processual transition largely intact, embedded now in assumptions about the social function of exchange rather than explicit theoretical commitments to cultural radiation.

Definition 1.7: The Default Prior as a Research Infrastructure Artifact

A theoretical position $\mathcal{H}$ becomes a *research infrastructure artifact* when its adoption as a default prior is sustained not primarily by the weight of evidence bearing on the specific question $\mathcal{H}$ addresses, but by the fact that the methods, training procedures, publication standards, and interpretive conventions of a research community are designed around $\mathcal{H}$'s truth, such that alternatives to $\mathcal{H}$ require not merely different conclusions but a different research apparatus to produce. Diffusionism in prehistoric archaeology exhibits this property: archaeologists trained in regional ceramic seriation, lithic typology, and culture-historical sequence-building are equipped to document the spread of ceramic types within a tradition but are not equipped, with the same toolkit, to test whether a superficially similar ceramic horizon in a geographically remote region arose independently or through contact. The test for independence requires methods — INAA fabric analysis, ancient DNA admixture testing, computational glottochronology — that are outside the core research infrastructure of most archaeology programmes and that require the kind of multi-disciplinary collaboration that institutional reward structures have, until recently, poorly incentivised. Changing the default prior from diffusionism to principled prior-weighting therefore requires not merely intellectual persuasion but institutional infrastructure investment.

1.8 The Comparative Scope of the Series: What It Covers and What It Does Not

The Network Horizon series as constituted covers four technologies across the Neolithic and Chalcolithic periods of the Old and New Worlds. This is a deliberate and principled scope restriction, not a statement that polycentric emergence is limited to these technologies, this period, or these worlds. The restriction reflects the availability of the full evidentiary toolkit: ancient DNA of adequate coverage and quality, INAA fabric databases of sufficient breadth, radiocarbon chronologies precise enough for the chronological feasibility tests of Definition 4.2, and proto-language reconstructions within the calibration horizon of Chapter 5's glottochronological analysis. Technologies for which one or more of these evidential streams is unavailable cannot be evaluated with the same rigour, and including them in the corpus on weaker grounds would dilute the epistemic standards of the series as a whole.

Proposition 1.3: Evidential Completeness as a Selection Criterion

For a technology $\mathcal{T}$ to be included in the Network Horizon calibration corpus (as opposed to the prospective prediction set of Chapter 6), it must satisfy a minimum evidential completeness criterion: the four-stream evidentiary framework of Definition 4.6 must be applicable to $\mathcal{T}$ with at least two of the four streams providing discriminating evidence (posterior likelihood ratio > 3 in either direction). Technologies for which the available evidence permits only the chronological stream to function, or for which no stream provides discrimination, are relegated to the prospective prediction set rather than the calibration corpus, because including non-discriminating cases in the calibration set would artificially inflate the apparent consistency of the model's fit to the evidence without providing any genuine test of the axiomatic structure.

The exclusion of post-Neolithic technologies from the calibration corpus reflects a further consideration beyond evidential completeness: the contact-network density condition of Axiom IV. By the Bronze Age in the Old World, and by the equivalent social-complexity threshold in the New World, regional interaction networks had in many areas crossed the Network Horizon ρ^* for most Neolithic technologies, placing them firmly in Regime II. The interesting empirical question for Bronze Age and later periods is not whether polycentric emergence occurred but whether the Regime I/II transition can be documented — whether the shift from structurally divergent to structurally convergent patterns in the material record can be dated to the period at which contact networks reached the technology-specific threshold. This is a research programme that follows from the series but is not itself part of it.

Corollary 1.2: The Regime I/II Transition as a Distinct Research Question

The question of when a given technology crossed its Network Horizon — when the contact-network density surpassed $\rho^*(\mathcal{T})$, causing the transition from polycentric divergence to diffusion-mediated convergence — is logically independent of the question of whether polycentric emergence occurred in the first place. Establishing that ceramic technology shows strong polycentric emergence in the period $20{,}000$–$8{,}000$ BP is compatible with ceramic traditions subsequently converging under contact-network densification in the Neolithic and Bronze Ages. The transition question requires a different methodological toolkit from the polycentricity question: it asks not whether early regional instances are independent but at what point they began to influence each other, and it requires the diachronic comparison of structural divergence measures discussed in Open Question 7.3. The present series establishes the baseline (the early polycentric phase) from which transition analyses can measure departure.

1.9 The Asymmetry of Burden of Proof Reconsidered

A recurrent objection to the framework developed in this series holds that the burden of proof should fall asymmetrically on the polycentric hypothesis rather than, as Chapter 1 argues, on whichever hypothesis the evidence least favours. The objection deserves a fuller hearing than it has so far received, because it is not merely a rhetorical move but reflects a substantive methodological position with a long pedigree in the philosophy of historical science.

The strongest form of the objection runs as follows: diffusion is a single causal mechanism (contact-mediated transmission) operating through well-understood channels (trade, migration, intermarriage, conquest), while independent invention requires that the same improbable conjunction of cognitive, ecological, and social conditions arise twice (or four times, or more) by chance. Since single-mechanism explanations are, all else equal, more parsimo-

nious than multiple-mechanism explanations, diffusion should retain a privileged status as the default hypothesis even after the four-stream evidence is in, with departures from that default requiring not merely evidence against diffusion but evidence whose strength scales with the number of independent origins postulated.

Note 1.4: Parsimony Arguments and Their Limits in Historical Inference

The appeal to parsimony in the diffusion-versus-independent-invention debate trades on an ambiguity between two distinct senses of simplicity. The first sense, familiar from philosophy of science as a tie-breaking criterion between empirically equivalent theories, favours the hypothesis that posits fewer causal mechanisms when the evidence underdetermines the choice between them. The second sense, relevant here, concerns the prior probability of a causal history given background knowledge about how often each type of event occurs. These two senses come apart precisely in cases like the present one: diffusion and polycentric emergence are not empirically equivalent (Axiom IV's contact-density condition makes them generate different predictions about exchange-network evidence), so the tie-breaking sense of parsimony does not apply. What remains is the prior-probability sense, and there the relevant question is not how many causal mechanisms are postulated but how probable each postulated mechanism is given what is independently known about cognitive convergence, ecological pressure, and technological constraint. A single low-probability mechanism (diffusion across a network with insufficient density, per Axiom IV) is not more parsimonious than several higher-probability mechanisms (independent invention under shared functional pressure, per Axiom III) merely because it is single.

This reframing does not settle the debate in favour of polycentric emergence; it removes parsimony as a free-standing argument for diffusion and returns the question to where the Network Horizon framework places it: an empirical matter to be resolved case by case through the four evidentiary streams. The asymmetric-burden objection, properly understood, is an argument that the prior probability of independent invention is low across all cases without exception — a much stronger and more falsifiable claim than the methodological parsimony argument it is often confused with. The axiomatic structure of Chapter 2 is, in effect, an attempt to specify precisely under what structural conditions that low-prior claim fails to hold.

Proposition 1.4: The Asymmetric-Burden Argument as an Empirical Claim in Disguise

The claim that diffusion should retain default status independent of case-specific evidence is logically equivalent to the claim that $\Pr(\mathcal{E}) \approx 0$ uniformly across all technologies and regions, where $\mathcal{E}$ is the sufficient-condition conjunction of Definition 1.3 (material redundancy, threshold accessibility, recurrent functional pressure, demographic isolation). This uniform-low-prior claim is itself an empirical hypothesis, testable by examining whether $\Pr(\mathcal{E})$ varies systematically across the technology–region pairs in the calibration corpus. If $\Sigma(\mathcal{T})$, the axiom-satisfaction composite of Definition 2.6, varies substantially across cases (as Chapter 3's cross-case synthesis demonstrates that it does, ranging from $\Sigma \approx 0.31$ for early ceramic technology in some regional pairings to $\Sigma \approx 0.89$ for horse domestication), the uniform-low-prior claim is falsified, and the asymmetric-burden argument loses its force as a general methodological position, however much it may continue to apply to specific technology–region pairs where Σ is in fact low.

1.10 Calibration Against Known Cases of Both Types

A framework that classifies all four technologies in its corpus as polycentric would be open to the objection that its axiomatic structure is tuned, consciously or not, to produce that result. This objection is addressed in part by the prospective prediction methodology of Chapter 6, but a complementary response is available within Chapter 1's epistemological framing: the model's axioms can be checked against historically documented cases of *bona fide* diffusion, where independent textual, epigraphic, or unambiguous artefactual evidence establishes contact-mediated transmission beyond reasonable dispute, to confirm that the axiomatic structure correctly classifies these as diffusion rather than spuriously inflating their polycentric scores.

Worked Example 1.3: Axiomatic Scoring of a Documented Diffusion Case

Consider the spread of iron smelting technology from a single innovation locus (most likely in Anatolia or the Caucasus, $c.\ 1{,}200$ BCE) across the Eastern Mediterranean and subsequently into temperate Europe, a case for which textual evidence (Hittite correspondence references to iron working), the chronological gradient of first appearances, and exchange-network density during the Late Bronze Age collapse and its aftermath jointly support contact-mediated transmission over independent invention. Scoring this case against the Network Horizon axioms: Axiom I (material redundancy) is satisfied, since iron ore is geologically ubiquitous; Axiom II (minimum viable threshold) is only marginally satisfied, since bloomery smelting requires a furnace temperature and a multi-stage reduction-and-forging sequence considerably more demanding than copper smelting, placing the technology near the upper bound of independently achievable complexity; Axiom III (recurrent functional pressure) is satisfied, since the functional advantages of iron tools over bronze were a generalised pressure across the Mediterranean and Near East; but Axiom IV (demographic isolation) is decisively *not* satisfied for the period and region in question, since exchange-network density in the Eastern Mediterranean circa 1200–900 BCE was well above the technology-specific Network Horizon $\rho^*(\text{iron})$, placing essentially all candidate regions firmly in Regime II. The overall axiomatic profile for iron smelting in this specific context (Eastern Mediterranean, 1200–900 BCE) is
emphweak, correctly and decisively favouring the diffusion hypothesis, consistent with the independent historical evidence. This negative calibration case is as important to the framework's credibility as the four positive cases: it demonstrates that the axiomatic structure is sensitive to the contact-density condition and does not simply assign high polycentric scores by default.

The iron smelting case also illustrates a structural point of some importance: it is precisely the case excluded from the calibration corpus proper (Proposition 1.3, evidential completeness) and relegated instead to a confirmatory role, because its primary discriminating evidence is textual and exchange-network rather than coming from the ancient-DNA or chronometric streams that define the corpus's evidential completeness criterion. Its inclusion here as a calibration check, rather than as a corpus member, reflects the methodological discipline the series aims to maintain: using a case for confirmation does not require including it in the formal corpus against which the axioms are validated, and conflating the two roles would blur the falsifiability structure established in Chapter 6.

Corollary 1.3: The Iron Smelting Case as a Negative Control

Because the iron smelting case scores low on Σ and is independently known to be a diffusion case, it functions as a negative control analogous to the function of negative

controls in experimental science: a case where the model is expected, on independent grounds, to output a low polycentric score, and does. The existence of at least one such negative control within the broader evidential landscape considered by the series (even though it falls outside the formal calibration corpus) partially addresses the tuning objection, since a model whose axioms were gerrymandered to always output high scores would not be expected to correctly classify a case it was not constructed to classify.

1.11 The Demarcation Problem: What Counts as a Technology for the Purposes of This Series

The preceding sections have spoken of "technologies" as if the unit of analysis were self-evident, but the demarcation of a technology — where one technology ends and another, related one begins — is itself a non-trivial methodological choice with consequences for how the axiomatic framework applies. Ceramic technology, for instance, could be treated as a single undifferentiated category, or decomposed into forming technique, firing technique, surface treatment, and vessel morphology as distinct sub-technologies each potentially subject to independent diffusion or independent invention.

Definition 1.8: Technology Individuation Criterion

For the purposes of the Network Horizon framework, a candidate unit $\mathcal{T}$ qualifies as a single technology if it satisfies functional unity (the unit accomplishes a single identifiable practical purpose, or a tightly bundled cluster of purposes that are not separately achievable), transmissibility unity (the knowledge or skill required to produce or use the unit is acquired as a single integrated package rather than as separable components that could be learned, transmitted, or invented independently), and archaeological visibility (the unit leaves a material signature that can be assessed against the four evidentiary streams independently of signatures left by related but distinct units). Ceramic forming technique and ceramic firing technique satisfy archaeological visibility independently — INAA fabric analysis bears primarily on the latter, vessel morphology and construction traces on the former — and arguably satisfy transmissibility unity independently as well, since ethnographic evidence documents potters acquiring forming technique and firing technique through partially separable apprenticeship pathways. This independence licenses their treatment, where the evidence permits, as distinct sub-technologies subject to separate axiomatic scoring, a refinement implicit in some of the more granular case analyses of Volume I but not made fully explicit until this discussion.

Note 1.5: The Risk of Arbitrary Decomposition

Granularity in technology individuation is not cost-free: decomposing a technology too finely risks generating spurious polycentric findings by treating what is, functionally and cognitively, a single integrated innovation as several independent ones, each scored separately and each individually satisfying the axioms more easily than the bundled whole would. A potter who develops a new firing technique typically does so in the context of, and partly motivated by, a specific forming tradition and a specific intended vessel function; treating firing technique as though it could be axiomatically scored in isolation from this context risks overstating the independence of what may, on closer inspection, be a single coherent technological development distributed across what the archaeological record happens to preserve as separable traces. The individuation criterion of Definition 1.8 guards against this risk via the transmissibility-unity condition, which

requires genuine evidence of separable acquisition pathways, not merely separable archaeological visibility, before a decomposition is licensed.

Worked Example 1.4: Applying the Individuation Criterion to Metallurgy

Copper metallurgy, as treated in Volume II, encompasses cold-working of native copper, smelting from oxide ores, and smelting from sulphide ores, each requiring progressively more demanding pyrotechnological control. Applying Definition 1.8: cold-working and oxide smelting satisfy functional unity only weakly with respect to each other, since cold-working accomplishes simple shaping of a soft material while smelting accomplishes a qualitatively different transformation (chemical reduction of ore to metal); they satisfy transmissibility unity weakly as well, since ethnographic and experimental evidence suggests smelting knowledge is not a simple extension of cold-working knowledge but requires distinct pyrotechnological understanding. The individuation criterion therefore licenses, and the series's actual practice reflects, treating native copper cold-working and ore smelting as distinct sub-technologies for axiomatic scoring purposes, consistent with the differential scoring already noted for native copper cold-working in the calibration discussion of Chapter 1. Oxide smelting and sulphide smelting, by contrast, share enough of the core pyrotechnological knowledge (furnace construction, temperature control, the basic reduction principle) that transmissibility unity is more plausibly satisfied between them, and the series treats sulphide smelting as a technical elaboration of oxide smelting rather than a fully distinct technology, consistent with standard archaeometallurgical practice.

1.12 The Status of Negative Cases: Technologies That Fail the Evidential Completeness Criterion

Proposition 1.3 established evidential completeness as a necessary condition for inclusion in the calibration corpus, relegating technologies that fail this criterion to the prospective prediction set of Chapter 6. This section examines several technologies that were considered for inclusion in the calibration corpus during the series' development but excluded on evidential-completeness grounds, both to clarify the corpus's boundaries and to illustrate the practical operation of Proposition 1.3's criterion.

Note 1.6: Excluded Candidate Technologies and the Reasons for Exclusion

Basketry and textile technology were considered as a potential fifth core case, given their plausible satisfaction of Axioms I–III (widely available plant-fibre raw materials, moderate cognitive complexity, recurrent functional pressure for containment and covering), but were excluded from the calibration corpus because organic preservation is sufficiently poor across most candidate comparison regions that the chronological stream cannot achieve the resolution needed for the feasibility tests of Definition 4.2, and because no analogue of INAA fabric analysis exists for perishable fibre technology, leaving only two of the four evidentiary streams (ancient DNA, via fibre-plant domestication genetics in some cases, and computational glottochronology) potentially functional, with neither providing strong independent discrimination given current data. Bow-and-arrow technology was similarly considered and excluded: while lithic microlith evidence provides good chronological resolution, no fabric-analogue stream exists, and the demographic-isolation evidence (Axiom IV) is confounded by the technology's early and widespread distribution, which makes regional comparison groups difficult to define with the pre-

cision the framework requires. Both technologies remain candidates for the prospective prediction set discussed in Chapter 6, where the weaker evidentiary requirements appropriate to genuinely prospective (rather than calibration) cases are more forgiving of partial evidence streams.

Proposition 1.5: Exclusion Is Not Disconfirmation

The exclusion of a candidate technology from the calibration corpus on evidential-completeness grounds carries no implication, positive or negative, about whether that technology in fact exhibits polycentric emergence or diffusion-mediated spread. Exclusion reflects a judgement about the current adequacy of available evidence relative to the framework's evidentiary standards, not a judgement about the underlying historical process. This distinction matters because it forecloses a possible misreading of the series' scope: the four core cases should not be read as having been selected because they are the four technologies for which polycentric emergence is true, with other technologies implicitly assumed to be diffusion cases by default. They are, rather, the four technologies for which the evidence is currently rich enough to support a confident determination either way, and the determination reached for each happens, in all four cases examined, to favour polycentric emergence.

1.13 Null Models and the Detection of Chance Convergence

The epistemological apparatus developed earlier in this chapter establishes that the burden of proof should fall on whichever hypothesis the evidence least favours, rather than defaulting asymmetrically to diffusion. This principle, however, presupposes that the evidence in question is capable of discriminating between hypotheses at all — that the observed pattern of chronological and structural convergence is not simply the kind of coincidence that would be expected even under a null model in which neither diffusion nor any structural predisposition toward independent invention is operative. Before the axiomatic apparatus of Chapter 2 is introduced, it is worth making explicit what such a null model would look like and why the four core cases of this series resist explanation by it.

Definition 1.9: The Null Model of Unstructured Technological Emergence

The *null model* $\mathcal{N}$ for technological emergence treats the appearance of a given technology in the archaeological record of any candidate region as a Poisson process with a constant per-region, per-century rate λ_0 that is independent of the region's specific ecological, demographic, or material circumstances — that is, a model in which technologies appear at random across the landscape with no structural predisposition favouring any particular region over any other, and with no mechanism (axiomatic or otherwise) linking the rate of appearance to the satisfaction of any set of structural preconditions. Under $\mathcal{N}$, an observed pattern of multiple roughly contemporaneous regional first-attestations is attributed to chance: a sufficiently large number of candidate regions observed over a sufficiently long period will, by the basic combinatorics of independent Poisson processes, occasionally produce clusters of near-simultaneous events even with no underlying structural cause.

The null model is a useful foil because it isolates exactly what the axiomatic framework of Chapter 2 is designed to explain away: not merely that independent invention occurred in multiple regions, but that it occurred in a pattern that is improbable under chance clustering and well-predicted by the specific structural conditions (material redundancy, threshold ac-

cessibility, recurrent functional pressure, demographic isolation, and functional convergence without structural homology) that Axioms I–V articulate. A finding of multiple regional origins that was equally well explained by $\mathcal{N}$ would not constitute evidence for the polycentric hypothesis in the sense intended by this series; it would constitute evidence only for the comparatively uninteresting claim that technological invention is not perfectly synchronised across all regions simultaneously, which no serious diffusionist position denies.

Theorem 1.9: Rejecting the Null Model Requires Structural, Not Merely Statistical, Evidence

Let k be the number of independent regional first-attestations of technology $\mathcal{T}$ observed within a window of width Δt, and let $\hat{\lambda}_0$ be the null-model rate estimated from the background frequency of technological innovations of comparable type across the full archaeological record. A purely statistical test of $\mathcal{N}$ — comparing the observed clustering of k events within Δt against the Poisson null — can reject $\mathcal{N}$ at conventional significance levels whenever k is large relative to $\hat{\lambda}_0 \Delta t$, but rejection of $\mathcal{N}$ on these grounds alone is compatible with diffusion as much as with polycentric emergence: a single-origin diffusion event followed by rapid spread also produces a cluster of regional first-attestations that would reject a constant-rate Poisson null. Distinguishing polycentric emergence from diffusion-driven clustering therefore requires the structural evidence captured by Axioms I–V (material independence, demographic isolation, structural divergence) in addition to, not instead of, the statistical rejection of $\mathcal{N}$. The chronological clustering pattern alone, however statistically improbable under $\mathcal{N}$, is symmetrically compatible with both substantive hypotheses and cannot itself discriminate between them.

This theorem clarifies a point of methodology that is easy to elide in informal presentations of the polycentric thesis: chronological proximity of regional first-attestations is necessary but radically insufficient evidence for polycentric emergence, since it is equally predicted by sufficiently rapid diffusion. The genuine discriminating evidence, throughout the four core cases examined in Volumes I–IV and synthesised in Chapter 3 below, comes from the structural evidence — material provenance, demographic isolation, structural divergence in production sequences — that the chronological pattern alone cannot provide. Readers who encounter popular presentations of polycentric-origins arguments that rest primarily on chronological clustering, without engaging the structural evidence, should treat such arguments with the same caution the present series applies to its own four cases.

Note 1.7: Why Chronological Clustering Is Nonetheless Reported Prominently

Given Theorem 1.9's caution that chronological clustering alone cannot discriminate between hypotheses, it might be asked why this series, and the broader literature it draws on, reports first-attestation dates and their clustering so prominently. The answer is that chronological data serves two distinct evidential functions that are easy to conflate. First, it provides the descriptive backbone against which the structural evidence is organised: knowing that ceramic technology appears in South China, the Russian Far East, the Sahara, and Amazonia within a multi-millennial window focuses the structural investigation on a specific, bounded set of cases rather than the unbounded space of all possible regional comparisons. Second, and more subtly, the chronological data enters the formal likelihood-ratio calculations of Chapter 4 and Chapter 5 not as freestanding evidence for polycentricity but as one of several inputs into the joint Bayesian assessment, where the chronological feasibility tests (Definition 4.2) ask the converse question — whether the observed gaps between regional first-attestations are too short to be explained by diffusion at archaeologically plausible velocities, which is a different and more demanding test

than mere clustering. The chronological evidence is reported prominently because it is necessary scaffolding and because, in its converse form, it does provide genuine discriminating power; it is not reported as though clustering alone settled the question.

Proposition 1.6: The Number of Candidate Regions Affects the Null Model's Predictive Power

The probability that the null model $\mathcal{N}$ produces, by chance, a cluster of k near-simultaneous regional first-attestations increases with the number of candidate regions n under consideration, for fixed per-region rate λ_0 and window width Δt, since more regions provide more opportunities for chance co-occurrence. This has a direct methodological consequence for the case-selection practices criticised in Definition 1.7's discussion of research infrastructure: if a researcher surveys a very large number of candidate regions and reports only the subset that exhibits clustering, the apparent improbability of the reported cluster under $\mathcal{N}$ is overstated, because the correct reference class for the statistical test is the full set of regions surveyed, not merely the subset selected for its clustering. The four core cases of this series were not selected from a larger survey of candidate regions screened for clustering; they were selected, as Chapter 1's demarcation discussion makes clear, on the prior grounds of evidential completeness (sufficient multi-source evidence to support a confident determination), with the clustering pattern itself — and its subsequent structural confirmation — emerging as a finding rather than a selection criterion.

Worked Example 1.5: Applying the Null-Model Test to a Hypothetical Five-Region Survey

Consider a hypothetical survey of $n = 40$ candidate world regions for evidence of a hypothetical technology $\mathcal{T}^{\dagger}$, conducted over a window $\Delta t = 5{,}000$ years, with a background rate $\hat{\lambda}_0 = 0.002$ per region per century estimated from comparable technologies. The expected number of independent first-attestations under $\mathcal{N}$ within any given 200-year sub-window is $n \cdot \hat{\lambda}_0 \cdot 2 = 40 \times 0.002 \times 2 = 0.16$, so observing $k = 4$ independent first-attestations within a single 200-year sub-window would be a 25-fold excess over the null expectation — a result that would, under a naive statistical test, appear highly significant. But because the survey examined $n = 40$ regions and the 5,000-year window contains 25 non-overlapping 200-year sub-windows, the relevant comparison is not whether any one sub-window shows excess clustering but whether the most-clustered of the 25 sub-windows does, after correction for multiple comparisons across sub-windows. Applying a Bonferroni-style correction for 25 comparisons raises the significance threshold substantially, and a result that appeared striking before correction may not survive it. This worked example illustrates concretely why Proposition 1.6's caution about reference-class selection is not a merely theoretical concern: it is a standard multiple-comparisons problem that any application of the null-model test to a broad regional survey must address explicitly, and that the four core cases of this series — selected on evidential-completeness grounds prior to any clustering analysis — are structurally protected from.

Corollary 1.4: The Null Model Is a Diagnostic Tool, Not a Competing Substantive Hypothesis

Unlike the diffusion hypothesis, which is a substantive claim about the mechanism of technological spread (contact-mediated transmission from a single origin), the null model $\mathcal{N}$ of Definition 1.9 is not a serious competitor for explaining any specific instance of multi-regional technological emergence; it is a diagnostic baseline against which both diffusion and polycentric emergence are evaluated for whether they offer genuine explanatory improvement over pure chance. No archaeologist seriously proposes that the

regional distribution of ceramic technology, copper metallurgy, horse domestication, or wheeled-vehicle technology in the prehistoric record reflects nothing more than unstructured chance; the live debate, which this series engages, is between diffusion and polycentric emergence as the two substantive explanations for non-chance patterning. The null model's role in the present chapter is to make explicit the standard against which both substantive hypotheses must clear before either is entitled to be treated as established, and to forestall a possible misreading of the series' conclusions as resting on clustering evidence that a careful null-model analysis would not in fact support.

1.14 The Role of Analogical Reasoning from Ethnographic and Historical Cases

A further methodological resource available to the framework, not yet discussed in the preceding sections of this chapter, is analogical reasoning from documented historical and ethnographic cases of technological spread and independent invention, where the contact networks and invention events are directly observed rather than archaeologically inferred. Such cases cannot substitute for the deep-time evidence the series' four core cases rely on, but they provide a valuable calibration check on the plausibility of the framework's parameter estimates.

Definition 1.10: Historically Documented Calibration Cases

A *historically documented calibration case* is an instance of technological spread or independent invention occurring within the period of written or otherwise directly documented history, for which the contact network, the timing of invention or adoption events, and the relevant material or cognitive prerequisites are independently verifiable from documentary sources rather than inferred solely from material remains. Such cases serve a calibration function distinct from the prospective application of Chapter 6: rather than testing the framework's predictions against future evidence, they test whether the framework's parameter estimates (transmission coefficients, threshold values, contact-density requirements) are consistent with directly observed historical analogues of the same general phenomena.

The independent invention of paper in Mesoamerica (*amatl*, attested from at least the Postclassic period) and in China (independently and considerably earlier, from the Han dynasty) is a useful historically adjacent calibration case: it is well attested that no plausible contact route connected these two paper-making traditions prior to the European colonial period, that both regions independently solved the structurally similar problem of producing a flat, durable, foldable writing or wrapping surface from beaten plant-fibre pulp, and that the specific production techniques diverged substantially between the two traditions, satisfying an Axiom-V-like divergence criterion. This case, while outside the prehistoric period the present series focuses on, lends qualitative support to the plausibility of the framework's general logic by providing an instance, closer to the documentary record, of the same convergence-without-contact pattern the axiomatic framework is designed to detect in deeper time.

Proposition 1.7: Historically Documented Cases Provide an Upper-Bound Check on Transmission Coefficients

Where a historically documented case involves directly attested contact between regions (rather than independent invention), the observed time between a technology's first appearance in one region and its documented arrival in a contacted region provides

a direct empirical estimate of the transmission coefficient $\alpha(\mathcal{T})$ for the relevant transport modality, unmediated by the inferential uncertainty that affects prehistoric velocity estimates. The documented spread of printing technology from China (woodblock printing, well attested by the ninth century CE) along Silk Road and subsequent maritime routes to the Islamic world and, much later and independently redeveloped with movable type, to Europe, provides one such check: the centuries-long delay and the substantial technical divergence between Chinese and European printing technology (movable type using different materials and mechanical principles in each case) is consistent with the relatively low transmission coefficients estimated for technologically demanding, high-tacitness innovations elsewhere in this volume (compare the wheel case's high-tacitness classification in the supplementary discussion of Chapter 4), lending plausibility to those estimates from a documented historical analogue.

Note 1.8: The Limits of Historical Analogy for Calibrating Deep-Time Parameters

The historical calibration cases discussed in this section should be treated as plausibility checks rather than as a basis for directly importing parameter values into the deep-time analysis of the four core cases, because the demographic, ecological, and social conditions of historically documented periods differ systematically from those of the Late Pleistocene and early Holocene periods relevant to ceramics, copper metallurgy, horse domestication, and the wheel. Population densities, settlement permanence, and the existence of specialised craft-production economies were all, in most historically documented cases, substantially greater than in the relevant prehistoric periods, which plausibly affects both the per-capita invention-attempt rate μ of Definition 2.5 and the transmission coefficients $\alpha(\mathcal{T})$ in ways that are not straightforwardly extrapolated backward. Historical analogy in this section therefore functions as a qualitative sanity check on the framework's general logic and the order of magnitude of its parameter estimates, not as a quantitative calibration source for the framework's deep-time applications.

Chapter 2

Axiomatic Foundations of the Network Horizon Model

2.1 The Need for Explicit Axiomatisation

The four preceding volumes each introduced formal definitions and findings within the context of a specific empirical case. The definitions were tailored to their cases: the loanword parsimony test of Volume IV, the diffusion-distance framework of Volume III, the cost-substitution model of Volume II. Taken individually, each formal apparatus is well-suited to its purpose. Taken together, they constitute an implicit theoretical framework that has never been stated as such — a set of assumptions, criteria, and inferential principles that are operative across all four volumes but distributed across four separate formal systems rather than unified in a single axiomatic structure.

This chapter provides that unification. The goal is not to replace the case-specific formal apparatus of the individual volumes but to identify the common assumptions underlying all of it, state them as explicit axioms, and derive from them the structural conditions for polycentric emergence that constitute the Network Horizon model in its general form. Explicit axiomatisation serves three purposes. First, it makes the theory's commitments transparent, enabling targeted criticism rather than diffuse disagreement. Second, it enables deductive extension: conclusions not derivable from any single volume's apparatus become derivable from the unified axiomatic structure. Third, it enables falsification: a theory stated as a set of axioms and derived propositions can be tested at the level of its axioms, not merely at the level of individual empirical findings.

2.2 Primitive Concepts

The Network Horizon model is built from five primitive concepts, which are taken as undefined and given meaning only through the axioms that govern their relations.

Definition 2.1: Primitive Concepts of the Network Horizon Model

The following five concepts are primitive (undefined) in the Network Horizon axiomatic system:

(i) **Technology** $\mathcal{T}$: a reproducible procedure for transforming available materials into artefacts that extend human adaptive capacity in a specifiable functional domain.

(ii) **Region** R: a geographically bounded area characterised by a specific suite of material resources, ecological conditions, and demographic history, treated as the unit

of analysis for invention and reception events.

(iii) **Invention event** $\iota(R, \mathcal{T}, t)$: the first occurrence, within region R at time t, of a reproducible implementation of technology $\mathcal{T}$ satisfying its functional specification, arising without external input from any region $R' \neq R$ that already possesses $\mathcal{T}$.

(iv) **Contact link** $\lambda(R_i, R_j, t)$: a directed or undirected relation between two regions at time t representing a channel through which material objects, persons, or symbolic information can be transmitted.

(v) **Sufficient condition set** $\mathcal{E}(R, \mathcal{T})$: the conjunction of material, ecological, demographic, and cognitive conditions in region R that, if jointly satisfied, render an invention event $\iota(R, \mathcal{T}, t)$ possible with non-negligible probability.

These five primitives are not empirical observations but theoretical posits; their application to specific cases requires operationalisation, which is the work of the case-specific formal apparatus developed in Volumes I–IV and reviewed systematically in Chapter 5.

2.3 The Axioms

Axiom I: Material Redundancy

For any technology $\mathcal{T}$ that is structurally prone to polycentric emergence, the raw materials required by its technological specification are available in geographically distributed, mutually independent sources, such that $\mathcal{E}(R, \mathcal{T})$ is satisfiable in at least $m \geq 2$ regions whose mutual contact-link density $\bar{\lambda}$ is below the Network Horizon $\rho^*(\mathcal{T})$:

$$\exists R_1, \ldots, R_m : \bigwedge_{i=1}^{m} \mathcal{E}(R_i, \mathcal{T}) \text{ satisfiable} \quad \wedge \quad \bar{\lambda}(\{R_1, \ldots, R_m\}) < \rho^*(\mathcal{T}).$$

Material redundancy is the necessary condition for polycentric emergence: without it, even if sufficient conditions are met locally, the absence of materials constrains invention to the subset of regions where the required inputs occur, reducing the effective number of independent invention sites.

Axiom II: Low Minimum Viable Threshold

For any technology $\mathcal{T}$ that is structurally prone to polycentric emergence, the minimum technical knowledge and skill required to produce a functional first instance of $\mathcal{T}$ — the *minimum viable threshold* $\theta(\mathcal{T})$ — is achievable by an individual or small group operating within the normal range of competences available in a Palaeolithic or early Neolithic forager community, without requiring prior knowledge of any more complex antecedent technology:

$$\theta(\mathcal{T}) \leq \theta_{\text{forager}},$$

where θ_{forager} denotes the upper bound on technical knowledge achievable by any competent adult member of a mobile forager group. Technologies with $\theta(\mathcal{T}) > \theta_{\text{forager}}$ (e.g., iron smelting, glass production) require cumulative knowledge transmission across generations or communities and are therefore structurally dependent on prior technological traditions in a way that makes independent invention at separated sites improbable.

Axiom III: Recurrent Functional Pressure

For any technology $\mathcal{T}$ that is structurally prone to polycentric emergence, the functional problem that $\mathcal{T}$ solves — the *adaptive pressure* $\mathcal{A}(\mathcal{T})$ — recurs independently in multiple geographically separated regions as a consequence of widely distributed ecological or demographic conditions rather than depending on a unique or locally specific configuration:

$$\Pr(\mathcal{A}(\mathcal{T}) \text{ present} \mid R) \text{ is high for all } R \text{ satisfying } \mathcal{E}(R, \mathcal{T}).$$

Recurrent functional pressure ensures that when material resources and technical capacity are locally available, the *motivation* to develop $\mathcal{T}$ is also present, converting the latent potential for invention into an active selection pressure. Without Axiom III, the conjunction of Axioms I and II would establish that independent invention is possible but not that it is likely; Axiom III provides the motivational component of the sufficient condition set.

Axiom IV: Demographic Isolation Below the Network Horizon

For any technology $\mathcal{T}$ at its time of first emergence, the candidate invention regions $\{R_1, \ldots, R_m\}$ are separated by contact densities below the Network Horizon $\rho^*(\mathcal{T})$, so that the expected diffusion transit time between any two candidate regions exceeds the expected inter-invention interval:

$$\tau_{\text{diff}}(\mathcal{T}, R_i, R_j) > \tau_{\text{inv}}(\mathcal{T}) \quad \forall i \neq j.$$

Axiom IV is the isolation condition: it specifies that the contact network at the time of emergence is insufficiently dense to carry $\mathcal{T}$ from its first invention site to the other candidate sites before those sites independently develop $\mathcal{T}$. This axiom is the one most directly testable by population genetic and archaeometric methods, since contact density is partially recoverable from ancient DNA admixture analysis and from the spatial patterning of material culture exchange.

Axiom V: Functional Convergence Without Structural Homology

For any technology $\mathcal{T}$ showing polycentric emergence, the independently developed regional instances $\mathcal{T}_1, \ldots, \mathcal{T}_m$ satisfy the functional specification of $\mathcal{T}$ (Axiom III's adaptive pressure $\mathcal{A}$) but need not be structurally identical: they may differ in raw material choices, production sequences, morphological forms, and decorative conventions while remaining functionally equivalent. Structural divergence between regional instances, in the absence of a common formal template, is positive evidence for independent invention; structural convergence beyond what is required by the functional specification is evidence for contact or common descent.

2.4 Derived Structural Conditions

From the five axioms, the following structural conditions are derivable as theorems rather than additional postulates.

Theorem 2.1: Joint Sufficiency of the Axioms for Polycentric Emergence

If a technology $\mathcal{T}$ satisfies Axioms I–IV simultaneously in a set of $m \geq 2$ regions $\{R_1, \ldots, R_m\}$, then the expected number of independent invention events $\mathbb{E}[|\{\iota(R_i, \mathcal{T}, t) :$

$i = 1, \ldots, m\}| > 1$, and the distribution of first-attestation dates $(\tau_1, \ldots, \tau_m)$ predicted by the polycentric model (P) assigns higher likelihood to the observed data than the diffusion model (D) whenever Axiom IV holds strictly.

Proof sketch: Under Axiom I, $\mathcal{E}(R_i, \mathcal{T})$ is satisfiable for all i. Under Axiom III, the adaptive pressure $\mathcal{A}$ is present in all R_i once $\mathcal{E}$ is satisfied, so the probability of an invention event $\iota(R_i, \mathcal{T}, t)$ in each region is non-negligible. Under Axiom IV, $\tau_{\text{diff}} > \tau_{\text{inv}}$, so the probability that a second invention event occurs before diffusion from the first event reaches the second region exceeds $\frac{1}{2}$ in expectation. Axiom II ensures that the invention events in different regions are genuinely independent (not requiring prior technological input from any common predecessor). Therefore the joint probability of m independent invention events strictly exceeds the probability of a single invention followed by $m - 1$ successful diffusion events under the contact constraints imposed by Axiom IV. *Q.E.D.*

Theorem 2.2: Axiom V as a Diagnostic — Structural Divergence Implies Independence

Under the polycentric model, the degree of structural divergence $\delta(\mathcal{T}_i, \mathcal{T}_j)$ between independently developed regional instances of $\mathcal{T}$ is bounded below by the structural variation introduced by independent solutions to the same functional problem:

$$\mathbb{E}[\delta(\mathcal{T}_i, \mathcal{T}_j) \mid \text{independent}] > \mathbb{E}[\delta(\mathcal{T}_i, \mathcal{T}_j) \mid \text{diffused}].$$

Observed structural divergence that exceeds the level expected under diffusion from a common template (where structural divergence is bounded by drift from the template) is therefore positive evidence for independence. The converse does not hold without qualification: functional convergence can generate structural similarity in the absence of contact, so structural similarity is not sufficient evidence for diffusion, though it is compatible with it.

Theorem 2.3: The Technology-Specificity of ρ^*

The Network Horizon $\rho^*(\mathcal{T})$ is a strictly decreasing function of $\theta(\mathcal{T})$ (Axiom II's minimum viable threshold) and a strictly increasing function of the geographic scope of material redundancy (Axiom I):

$$\frac{\partial \rho^*}{\partial \theta} < 0, \qquad \frac{\partial \rho^*}{\partial |\text{material range}|} > 0.$$

Technologies with low θ and wide material redundancy have high ρ^*: they remain below their Network Horizon even in relatively dense contact networks, and therefore tend to show polycentric emergence across a wide range of prehistoric settings. Technologies with high θ and narrow material redundancy have low ρ^*: they are above their Network Horizon even in sparse contact networks, and tend to diffuse from single invention sites. This theorem provides the structural basis for predicting which future technologies (not yet subject to polycentric analysis) are likely to show polycentric signatures and which are likely to show single-origin diffusion.

2.5 The Parameter Space of the Model

The five axioms and three theorems above define a two-dimensional parameter space within which any technology can be located. The first dimension is $\theta(\mathcal{T})/\theta_{\text{forager}}$, the ratio of the technology's minimum viable threshold to the forager competence ceiling (Axiom II). The second

dimension is $\bar{\lambda}/\rho^*(\mathcal{T})$, the ratio of the observed inter-regional contact density to the technology's Network Horizon (Axiom IV). Four regions of this parameter space are theoretically distinct.

Definition 2.2: The Four Regimes of the Parameter Space

- **Regime I** ($\theta/\theta_F < 1$, $\bar{\lambda}/\rho^* < 1$): Low threshold, sparse contact. *Predicted outcome*: strong polycentric emergence. Multiple independent invention events expected; structural divergence between regional instances high.
- **Regime II** ($\theta/\theta_F < 1$, $\bar{\lambda}/\rho^* > 1$): Low threshold, dense contact. *Predicted outcome*: diffusion dominates despite low threshold; independent re-invention possible but pre-empted by rapid spread from first invention site.
- **Regime III** ($\theta/\theta_F > 1$, $\bar{\lambda}/\rho^* < 1$): High threshold, sparse contact. *Predicted outcome*: neither independent re-invention nor diffusion is likely; technology is confined to the region where a sufficiently advanced knowledge base pre-exists.
- **Regime IV** ($\theta/\theta_F > 1$, $\bar{\lambda}/\rho^* > 1$): High threshold, dense contact. *Predicted outcome*: single-origin diffusion once the high threshold is surmounted; technology spreads rapidly along the dense contact network from its single invention site.

Technologies in Regime I are the primary subject of this series. The four cases of Volumes I–IV are all argued to fall in Regime I at their times of emergence.

Note 2.1: The Regime I/II Boundary as a Historical Transition

A technology does not remain in the same regime across time. As regional contact networks densify through population growth, sedentism, and the accumulation of exchange relationships, $\bar{\lambda}$ increases while $\rho^*(\mathcal{T})$ remains fixed (it is a property of the technology, not of the social context). A technology that began in Regime I may therefore transition into Regime II as its social context evolves, after which any further independent re-invention events are pre-empted by diffusion from established traditions. This historical transition from Regime I to Regime II is observable in the archaeological record as a shift from structural divergence between regional instances to structural convergence — the signature of a technology that has entered the diffusion-dominated phase of its spread. Several of the regional ceramic traditions documented in Volume IV show this transition over their late prehistoric sequences.

Note 2.1b: The Epistemic Status of σ, θ, and ρ^* — Ordinal Expert Judgements, Not Measurements

Before these parameters are used in the case applications of Chapter 3 and the predictions of Chapter 6, their epistemic status needs to be stated without the hedge of surrounding formalism. The axiom-satisfaction scores $\sigma_1, \ldots, \sigma_4$ reported throughout this volume (e.g. $\sigma_1 \approx 0.95$ for agriculture's material redundancy) are not measurements in the sense that a radiocarbon date or an isotope ratio is a measurement. They are structured expert elicitations: a domain specialist's judgement of how strongly a given body of qualitative evidence satisfies a given axiom, expressed on a $[0, 1]$ scale for combinability, but without an underlying metric whose units the second decimal place actually tracks. The same caveat applies to $\theta(\mathcal{T})/\theta_{\text{forager}}$ and to $\rho^*(\mathcal{T})$ itself: neither has been estimated from an independent dataset using a stated measurement procedure with reportable error bars; both are derived from the converging judgement of the case-specific analyses in

Volumes I–IV, organised post hoc into the numerical form used here.

This matters for how every subsequent numerical claim in this volume should be read. A reported value such as $\Sigma(\text{ceramics}) = 0.69$ should be understood as shorthand for "domain specialists judge ceramic technology to satisfy the axiomatic conditions strongly, on most or all of the four dimensions, with no serious disqualifying counter-evidence on any one axiom" — not as a quantity with the precision its two significant figures suggest. Where this volume reports numerical scores to two decimal places, the practice should be read as a notational convenience for combining four ordinal judgements consistently, not as a claim of measurement precision the underlying evidence does not support. Readers should treat differences of 0.05–0.10 between two scores as not meaningfully distinguishable; only differences large enough to move a case across a regime boundary (Definition 2.2) should be treated as substantively informative. Genuine quantitative calibration — scores derived from a stated, independently reproducible measurement procedure with error bars — is identified in Chapter 7 as the framework's single most important unmet methodological need, not as a refinement already achieved.

2.6 Relationship to Existing Theoretical Frameworks

The Network Horizon model is not proposed in a theoretical vacuum. Three existing frameworks have addressed related questions and it is important to specify where the Network Horizon model agrees, extends, or disagrees with each.

Case Study 2.1: World Systems Theory and Its Limits

World Systems Theory (Wallerstein 1974; Chase-Dunn and Hall 1991 for prehistoric applications) models technological diffusion as a function of core-periphery relationships in inter-regional exchange networks. It correctly identifies the importance of network structure for technology transmission and provides a framework for understanding why some technologies diffuse rapidly (those controlled by core regions with dense exchange links) while others diffuse slowly or not at all (those confined to peripheries). Its limitation, for the purposes of the Network Horizon model, is that it assumes a single global network within which all regions are embedded at varying distances from the core. It has no mechanism for representing regions that are genuinely outside the relevant network — below the contact-link threshold for receiving any given technology — and therefore cannot account for cases where local invention is structurally predicted rather than merely permitted. The Network Horizon model's Axiom IV, which makes the contact density threshold explicit and technology-specific, addresses this gap.

Case Study 2.2: Niche Construction Theory

Niche construction theory (Odling-Smee, Laland, and Feldman 2003; Zeder 2012 for domestication) models technological change as a feedback between human activity and the selective environment: technologies modify the ecological niche, which in turn modifies the selective pressures on further technological development. It provides a powerful framework for understanding why multiple independent domestication events occurred in ecologically similar but geographically separated refugial zones — a case structurally analogous to the ceramic polycentric emergence documented in Volume IV. The Network Horizon model's Axiom III (recurrent functional pressure) is compatible with the niche construction framework's account of how similar ecological contexts generate similar selective pressures on technology. Where the frameworks diverge is in their treatment of the

contact network: niche construction theory is primarily ecological and does not formalise the role of inter-regional contact density in determining whether independent inventions occur before diffusion from prior events.

Case Study 2.3: The *Chaîne Opératoire* Tradition

The *chaîne opératoire* framework (Mauss 1935; Lemonnier 1992; Roux 2019) analyses technology as structured sequences of gestures, knowledge states, and material transformations, focusing on the transmission of technical knowledge through embodied practice rather than on the spatial spread of artefact types. It provides the most fine-grained account of what is actually transmitted when a technology diffuses — not merely the artefact form but the specific sequence of actions required to produce it — and its diagnostic criterion (the sharing of non-functional technical choices, les *faits techniques*) is among the most powerful tools for distinguishing shared tradition from independent convergence. The Network Horizon model's Axiom V (functional convergence without structural homology) is directly compatible with the *chaîne opératoire* approach: structural homology beyond what functional requirements dictate is precisely what the *chaîne opératoire* framework identifies as the signature of transmitted tradition. The Network Horizon model's contribution is to embed this diagnostic within a formal framework that also accounts for the network conditions under which transmission is possible.

Case Study 2.4: The Childe–Renfrew Debate as the Direct Disciplinary Antecedent

The most direct antecedent to the Network Horizon framework's central question is not a general theoretical school but a specific, decades-long disciplinary dispute: V. Gordon Childe's diffusionist account of European prehistory, in which innovations including metallurgy and (in his earlier work) megalithic monument-building were held to radiate from Near Eastern centres, against Colin Renfrew's revisionist case (*Before Civilization*, 1973) that radiocarbon dating undermined the chronological sequencing on which Childe's diffusion routes depended and that several European technological developments, megalith construction prominent among them, were better explained as local innovations. The Childe–Renfrew dispute is instructive for the present framework in a way the comparative cases above are not, because it was fought and substantially settled on the same evidentiary terrain — chronometric dating — that Axiom IV's contact-density and transit-time arguments depend on here. Renfrew's case did not rest on a general theory of polycentric emergence; it rested on showing, region by region, that the dates required by Childe's diffusion model were not supported once corrected radiocarbon chronologies were available. This is a narrower and more empirically disciplined form of argument than the axiomatic apparatus developed in this chapter, and the difference is worth stating plainly: Renfrew falsified specific diffusion claims by re-dating; the Network Horizon model proposes a general structural account of *why* re-dating tends to favour independent invention for some classes of technology and not others. The model's claim to add value beyond the Childe–Renfrew precedent rests entirely on whether the axiomatic structure of this chapter generates correct predictions for cases Renfrew's chronometric method alone could not have anticipated — a claim assessed directly in Chapter 6.

Note 2.2: What the Childe–Renfrew Precedent Implies About This Framework's Originality

A reader familiar with the history of archaeological theory may reasonably ask what the Network Horizon model contributes that the post-Renfrew consensus already established. The honest answer is qualified. The post-Renfrew consensus already accepts,

case by case, that many specific instances of apparent diffusion do not survive corrected chronology, and that independent local development is frequently the better-supported explanation. What the post-Renfrew literature has not generally done is state, in advance of examining a specific case, the structural conditions under which independent development *should be expected* versus the conditions under which diffusion remains the more probable account — it has proceeded case by case, largely without a general predictive apparatus. The Network Horizon model's claim to originality, such as it is, lies in that predictive apparatus and in the falsifiability programme of Chapter 6, not in the underlying empirical conclusion that independent invention occurred in specific cases, which in several instances (megalith construction prominently) Renfrew and others had already established by other means decades before this series was conceived.

2.7 A Note on the Wheel as a Limiting Case

The four cases examined in Volumes I–IV do not sit uniformly within the axiomatic framework. The wheel (Volume II) is the case where the polycentric emergence argument is weakest, and it is useful to understand why in axiomatic terms. The wheel satisfies Axiom I (wood and axle materials are widely available), Axiom III (transport efficiency pressure is recurrent), and Axiom V (regional formal variation in early wheel-axle assemblages is attested). It partially satisfies Axiom II: the minimum viable threshold for a functional cart wheel is not extreme, but it requires a level of woodworking precision — specifically, the ability to produce a sufficiently round wheel and a sufficiently smooth axle fitting — that is at the upper end of the forager competence range. The critical difficulty is Axiom IV: the current archaeological evidence places the earliest wheel-axle assemblages in a relatively narrow chronological and geographic window (Pontic-Caspian steppe, $c.\ 3500$ BCE), with secondary attestations that are consistent with rapid diffusion from this region along well-documented Bronze Age exchange networks. For the wheel, $\bar{\lambda}/\rho^*$ may be close to or slightly above 1 at the relevant period, placing it near the Regime I/II boundary rather than firmly in Regime I.

Note 2.2: The Wheel as a Regime I/II Boundary Case

The Network Horizon model does not require all four cases to sit identically within the parameter space. It predicts that cases in Regime I show strong polycentric signatures, cases near the Regime I/II boundary show ambiguous or mixed signatures, and cases in Regime II show single-origin diffusion signatures. The wheel's position near the boundary is therefore not an anomaly for the model but a predicted intermediate case. The model would be in difficulty if all four cases showed the same degree of polycentric evidence regardless of their axiomatic properties; the wheel's relative weakness as a polycentric case, on the axiomatic analysis, is actually a point in the model's favour as a discriminating theoretical framework.

2.8 Summary of the Axiomatic Structure

The Network Horizon model rests on five axioms (Material Redundancy, Low Minimum Viable Threshold, Recurrent Functional Pressure, Demographic Isolation Below the Network Horizon, and Functional Convergence Without Structural Homology) and three derived theorems (Joint Sufficiency, Structural Divergence as Independence Diagnostic, Technology-Specificity of the Network Horizon). The axioms are not individually sufficient for polycentric emergence: any single axiom can be satisfied while the others are violated, in which case diffu-

sion or technological stasis is predicted. Their joint satisfaction is what constitutes structural proneness to polycentric emergence in the sense of Definition 1.3.

The following chapter applies this axiomatic framework retrospectively to the four cases of Volumes I–IV, assessing the degree to which each case satisfies each axiom and mapping each technology onto the parameter space of Definition 2.2.

2.9 The Ontological Status of Axioms in Archaeological Theory

The five axioms of the Network Horizon model (Sections 2.3–2.7) are not empirical generalisations derived by induction from the four cases of Volumes I–IV. They are structural conditions that a technology must satisfy for polycentric emergence to be structurally predicted, stated at a level of generality that makes their content independent of any specific empirical case. This distinction — between an axiom as a condition and an empirical generalisation as a pattern — matters for the epistemology of the framework, and it is worth making explicit before the axioms are applied retrospectively in Chapter 3.

An empirical generalisation of the form "all technologies examined in this series satisfy Axiom III" would be trivially true if the cases were selected precisely because they exhibit the pattern — a classic instance of confirmation bias dressed in axiomatic clothing. The Network Horizon axioms avoid this objection because their content can be assessed, for any technology including ones not in the calibration set, by reference to evidence that is independent of the chronological convergence pattern that motivates the polycentric hypothesis. Axiom I (material redundancy) depends on geology and palaeobotany, not on ceramic chronology. Axiom II (minimum viable threshold) depends on experimental archaeology and comparative technical complexity studies. Axiom III (recurrent functional pressure) depends on palaeoecology and subsistence analysis. Axiom IV (demographic isolation) depends on population genetics and trade-network reconstruction. Axiom V (functional convergence without structural homology) depends on archaeometric analysis of the artefacts themselves.

Theorem 2.5: Independence of Axiom Assessment from Convergence Evidence

Let $\mathbf{E}_{\text{conv}}$ denote the evidence bearing directly on the chronological convergence pattern — the first-attestation dates $(\tau_1, \dots, \tau_n)$ and their spatial distribution — and let $\mathbf{E}_k$ denote the evidence bearing on Axiom A_k, $k = 1, \dots, 5$. The axiom-satisfaction score $\sigma_k(\mathcal{T}) = \Pr(A_k \text{ satisfied} \mid \mathbf{E}_k(\mathcal{T}))$ is *independent* of $\mathbf{E}_{\text{conv}}$ if and only if:

$$\Pr(A_k \text{ satisfied} \mid \mathbf{E}_k(\mathcal{T}), \mathbf{E}_{\text{conv}}) = \Pr(A_k \text{ satisfied} \mid \mathbf{E}_k(\mathcal{T})).$$

This conditional independence holds for all five axioms when the evidential sources for each axiom are distinct from the sources providing the convergence evidence: geological surveys for Axiom I, experimental replication studies for Axiom II, palaeoecological reconstruction for Axiom III, ancient DNA and exchange-network data for Axiom IV, and archaeometric fabric analysis for Axiom V. Violations of this independence arise when, for example, the same radiocarbon dates are used both to establish the chronological pattern and to assess Axiom IV's contact-density condition — a circularity risk discussed in Section 5.6. When the evidence is properly partitioned, the axiom-satisfaction scores provide genuinely independent constraints on the polycentric model, and their product $\Sigma(\mathcal{T})$ is a non-trivial prediction of the model's fit to the convergence evidence.

This independence property is what gives the composite polycentric score $\Sigma(\mathcal{T})$ its predictive significance. A score computed from evidence that is not independent of the convergence pattern would simply be a restatement of the convergence evidence in axiomatic notation,

with no predictive power for new cases. The independence of the axiomatic evidence streams ensures that Σ is a genuine predictor — derived from structural properties of the technology and its context that are assessable before the convergence pattern is examined — and that its retroactive consistency with the observed convergence (Theorem 2.1) is a non-trivial result.

2.10 The Parameter ρ^* and Its Estimation

Definition 1.4 introduced the Network Horizon $\rho^*(\mathcal{T})$ as the critical contact density above which diffusion of $\mathcal{T}$ outpaces independent re-invention. Theorem 2.3 established that ρ^* is a strictly decreasing function of the minimum viable threshold $\theta(\mathcal{T})$ and a strictly increasing function of the geographic scope of material redundancy. These results characterise ρ^* qualitatively but do not prescribe a method for estimating it numerically in any specific case. This section addresses the estimation problem.

Definition 2.3: Operational Estimate of ρ^*

For a technology $\mathcal{T}$ with observed first-attestation dates $(\tau_1, \dots, \tau_n)$ in regions $\mathcal{R}$, the operational estimate of the Network Horizon is:

$$\hat{\rho}^*(\mathcal{T}) = \frac{v_{\max}(\bar{\lambda})}{d_{\min}(\mathcal{R})} \cdot \tau_{\text{inv}}(\mathcal{T}),$$

where $d_{\min}(\mathcal{R}) = \min_{i \neq j} d(R_i, R_j)$ is the minimum pairwise distance between candidate invention regions along the least-cost contact path, $v_{\max}(\bar{\lambda})$ is the maximum plausible diffusion velocity at contact density $\bar{\lambda}$ (estimated from ethnographic and historical analogues), and $\tau_{\text{inv}}(\mathcal{T})$ is the expected inter-invention interval estimated from the axiom-satisfaction scores via:

$$\tau_{\text{inv}}(\mathcal{T}) \approx \frac{1}{\mu(\mathcal{T}) \cdot P(\mathcal{E})}$$

where $\mu(\mathcal{T})$ is the per-capita invention probability per unit time conditional on the sufficient condition set $\mathcal{E}$ being satisfied, and $P(\mathcal{E})$ is the probability that $\mathcal{E}$ is jointly satisfied in a given region at a given time, estimated from the axiom-satisfaction scores.

The estimation procedure implicit in Definition 2.3 has a circular component that must be explicitly acknowledged: $v_{\max}(\bar{\lambda})$ depends on $\bar{\lambda}$, which is the quantity whose relation to ρ^* is under evaluation. This circularity is broken by the independence of the axiom-assessment evidence ($\mathbf{E}_k$) from the convergence evidence ($\mathbf{E}_{\text{conv}}$), established in Theorem 2.5: the contact density $\bar{\lambda}$ is estimated from population genetic and exchange-network data that is logically prior to the technology-specific convergence pattern, and the axiom-satisfaction scores that determine τ_{inv} are estimated from evidence independent of $\bar{\lambda}$. The resulting estimate of $\hat{\rho}^*$ is not, therefore, a function of the convergence pattern itself, even though it shares some evidential sources with the contact-density estimate.

Proposition 2.2: The Sensitivity of $\hat{\rho}^*$ to Input Parameter Uncertainty

Under a log-normal error model for the input parameters $v_{\max}$, $d_{\min}$, and τ_{inv}, the resulting uncertainty in $\hat{\rho}^*$ propagates as:

$$\frac{\sigma_{\hat{\rho}^*}}{\hat{\rho}^*} \approx \sqrt{\left(\frac{\sigma_{v_{\max}}}{v_{\max}}\right)^2 + \left(\frac{\sigma_{d_{\min}}}{d_{\min}}\right)^2 + \left(\frac{\sigma_{\tau_{\text{inv}}}}{\tau_{\text{inv}}}\right)^2}.$$

For the ceramics case, where $v_{\max}$ is constrained by LGM demographic data to within a

factor of 3, $d_{\min}$ is geographically fixed within $\pm 10\%$, and τ_{inv} is estimated from the axiom scores to within a factor of 5, the relative uncertainty in $\hat{\rho}^*$ is approximately 60% — large in absolute terms but small relative to the three-order-of-magnitude difference between the estimated $\bar{\lambda}$ and $\hat{\rho}^*$ for the ceramics case. The ceramics classification as firmly in Regime I is therefore robust to the full range of input parameter uncertainty. The wheel classification, where $\bar{\lambda}/\hat{\rho}^*$ is estimated to be near 1, is correspondingly sensitive to parameter uncertainty, which is reflected in the wide 95% HPD interval of Finding 4.3.

2.11 Extensions of the Axiomatic Framework

The five axioms and three theorems of the Network Horizon model constitute a minimal axiomatic system — one from which the core structural conditions for polycentric emergence are derivable, without additional postulates. Whether additional axioms should be added to handle cases that fall in the model's margins — the wheel, iron smelting, and the other borderline technologies discussed in Chapter 6 — is a question that the current series does not definitively answer but that the framework makes precisely askable.

Definition 2.4: A Candidate Sixth Axiom — Social Technology Receptivity

Some technologies require not only the material and cognitive conditions of Axioms I–III but also a specific configuration of social institutions that permits the technology's integration into the regional economy and that rewards its development and refinement. Define the *social technology receptivity* $\mathcal{R}_S(\mathcal{T}, R)$ as the degree to which the social organisation of region R provides the institutional context — property rights over innovations, apprenticeship or knowledge-transmission structures, markets or prestige systems rewarding technological proficiency — necessary for technology $\mathcal{T}$ to be adopted, refined, and maintained across multiple generations once initially invented. A candidate Axiom VI would require high $\mathcal{R}_S(\mathcal{T}, R)$ in all candidate regions for polycentric emergence to be predicted. For the four core cases of this series, $\mathcal{R}_S$ is not the binding constraint: ceramic production, horse management, copper-working, and wheel-making can each be maintained within the social institutions of a small forager or early agropastoral community. But for technologies that require more complex institutional frameworks — writing systems, standardised coinage, legal codes — Axiom VI would identify the limiting factor that Axioms I–V cannot capture.

The candidate Axiom VI illustrates a broader point about the relationship between the Network Horizon model and theories of social evolution. The model's current axioms focus on material, cognitive, and demographic conditions; they abstract away from the internal social dynamics of the invention-and-adoption process within any given region. This abstraction is appropriate for the four core cases, where social structure is not the binding constraint on polycentric emergence, but it becomes an oversimplification when applied to technologies whose emergence depends critically on institutional prerequisites. A full theory of prehistoric technological emergence would require the integration of the Network Horizon model's material and demographic axes with a theory of institutional evolution — a project well beyond the scope of this series but delineated by it.

Note 2.3: The Relationship to Evolutionary Game Theory

The question of how social institutions enabling technology adoption evolve is naturally modelled within evolutionary game theory (Maynard Smith 1982; Skyrms 1996).

In a population of agents who can either adopt a new technology or free-ride on the adoptions of others, the adoption equilibrium depends on the payoff structure of the technology, the size and connectivity of the population, and the presence or absence of punishment mechanisms for free-riding. The Network Horizon model's Axiom IV (demographic isolation) has a natural interpretation in this framework: sparse contact networks reduce the scope for free-riding on distant adopters' knowledge, which in turn increases the selective pressure on local adoption and independent refinement. Low $\bar{\lambda}$ therefore promotes not only multiple independent invention events but also the development of local institutional frameworks for technology maintenance, because the option of importing knowledge from connected regions is foreclosed. This connection between the demographic isolation condition and the evolutionary dynamics of technology adoption is a theoretical thread that future work could develop into a formal link between the Network Horizon model and the evolutionary game-theoretic literature on cultural evolution.

2.12 Worked Derivation: Computing τ_{inv} from Axiom-Level Parameters

The inter-invention interval $\tau_{\text{inv}}(\mathcal{T})$ is a central parameter of the Network Horizon model that appears in the Network Horizon definition (Definition 1.4), the axiomatic sufficiency theorem (Theorem 2.1), and the diffusion boundary ratio (Definition 4.7). Its estimation from the axiom-level parameters is not straightforward: the five axiom-satisfaction scores are probabilities that the structural conditions for invention are met, not direct estimates of the time between independent invention events. This section provides a worked derivation of τ_{inv} from the axiomatic parameters, using the ceramics case as an illustration.

The fundamental relationship is between the per-region probability of an invention event per unit time, $p_{\text{inv}}(R, \mathcal{T}, t)$, and the axiom-satisfaction scores. Under the assumption that an invention event occurs in region R at time t if and only if all five axiomatic conditions are jointly met and the region's population is large enough to produce an inventor who attempts the relevant combination of materials and techniques, the per-region invention probability is:

Definition 2.5: Per-Region Invention Probability

The per-region probability of an invention event $\iota(R, \mathcal{T}, t)$ per unit time (here, per century) is:

$$p_{\text{inv}}(R, \mathcal{T}, t) = N(R, t) \cdot \mu \cdot \prod_{k=1}^{5} \sigma_k(\mathcal{T}) \cdot f(R, \mathcal{T}, t),$$

where $N(R, t)$ is the effective population size of region R at time t, μ is the per-capita per-century probability of a cognitively mature individual in a community satisfying all axiomatic conditions attempting the specific combination of actions that constitutes the technology's invention, $\prod_{k=1}^{5} \sigma_k(\mathcal{T})$ is the probability that all axiomatic conditions are jointly met in region R (the composite polycentric score $\Sigma(\mathcal{T})$), and $f(R, \mathcal{T}, t)$ is a region-specific adjustment factor reflecting local ecological and demographic conditions that enhance or reduce the invention probability beyond what the axiom scores capture.

For the ceramics case, the parameter values are estimated as follows. The effective population size of the South Chinese candidate region during the LGM is estimated from palaeodemographic modelling (Bocquet-Appel et al. 2012) at $N \approx 2{,}000$–$5{,}000$ individuals. The per-capita attempt probability μ is estimated from the experimental archaeology literature:

studies of naive learners' encounter with clay-working find that spontaneous experimentation with plastic clay in communities that have access to fire leads to functional firing attempts within 1–5 generations in approximately 1%–5% of communities per century, implying $\mu \approx 10^{-4}$–5×10^{-4} per capita per century. With the composite axiomatic score for ceramics (approximately 0.69 under the product rule, as the shorthand notation for the uniformly strong axiom profile of Finding 3.1, per Note 2.1b) and a population of $3{,}000$:

$$p_{\text{inv}}(\text{South China}, \text{ceramic}, t_{\text{LGM}}) \approx 3{,}000 \cdot 3 \times 10^{-4} \cdot 0.69 \approx 0.62 \text{ per century},$$

implying an expected inter-invention interval of approximately $\tau_{\text{inv}} \approx 1/0.62 \approx 1.6$ centuries for the South Chinese region alone. Given five candidate regions, the expected time between any two independent invention events across the global ceramic candidate set is approximately $\tau_{\text{inv}}/5 \approx 30$ years — consistent with the interpretation that multiple independent invention events occur rapidly once the demographic and material conditions of the LGM-to-Holocene transition are crossed.

Proposition 2.3: Sensitivity of τ_{inv} to Population Size Assumptions

The expected inter-invention interval $\tau_{\text{inv}}(\mathcal{T})$ is inversely proportional to the total effective population across all candidate regions satisfying the axiomatic conditions:

$$\tau_{\text{inv}}(\mathcal{T}) = \frac{1}{\mu \cdot \Sigma(\mathcal{T}) \cdot \sum_{i=1}^{m} N(R_i, t)}.$$

Uncertainty in the palaeodemographic estimates for the LGM translates directly into uncertainty in τ_{inv}: a factor-of-three uncertainty in N produces a factor-of-three uncertainty in τ_{inv}. For the ceramics case, this uncertainty is irrelevant to the model's qualitative conclusions because τ_{inv} is hundreds of years while τ_{diff} is thousands to tens of thousands of years for the relevant region pairs — the factor-of-three uncertainty does not bring the two quantities into competition. For the wheel case, where $\tau_{\text{diff}}/\tau_{\text{inv}}$ is estimated to be close to 1, the palaeodemographic uncertainty is more consequential and should be explicitly incorporated in the boundary ratio distribution of Definition 4.7.

2.13 The Network Horizon in Formal Analogy to Percolation Theory

The Network Horizon condition $\bar{\lambda} = \rho^*(\mathcal{T})$ has a formal analogy to the percolation threshold in statistical physics. In bond percolation on a graph, the percolation threshold p_c is the critical bond occupation probability above which an infinite connected cluster spans the graph, enabling long-range connectivity. Below p_c, the graph consists of finite, disconnected clusters; above p_c, a connected pathway crosses the entire system. The Network Horizon ρ^* plays an analogous role: below it, the contact network does not support sustained long-range technology diffusion (the relevant "cluster" of connected sites is smaller than the distance between candidate invention sites); above it, the network supports rapid diffusion across the full candidate region set.

Theorem 2.6: Percolation Analogy for the Network Horizon

On a random geometric graph G with n nodes uniformly distributed in a two-dimensional domain of area A and edges between all pairs within radius r (proportional

to $\bar{\lambda}$), the percolation threshold ρ_c for an infinite connected component scales as:

$$\rho_c \sim \frac{\pi r_c^2}{A/n} = \pi \lambda_{\text{critical}} n/A,$$

where r_c is the critical radius. The Network Horizon $\rho^*(\mathcal{T})$ in the prehistoric contact-network setting maps to ρ_c in the percolation analogy, with the correspondence:

$$\begin{aligned} \bar{\lambda} &\leftrightarrow \text{bond occupation probability } p \\ \rho^*(\mathcal{T}) &\leftrightarrow \text{percolation threshold } p_c \\ \tau_{\text{diff}} &\leftrightarrow \text{cluster correlation length}^{-1} \\ \tau_{\text{inv}} &\leftrightarrow \text{spontaneous activation rate} \end{aligned}$$

The analogy is not exact — prehistoric contact networks are heterogeneous, spatially structured, and temporally dynamic in ways that simple random geometric graphs are not — but it provides intuition for why the Network Horizon transition has the character of a phase transition rather than a smooth crossover: the shift from Regime I to Regime II as $\bar{\lambda}$ crosses ρ^* involves the sudden appearance of a connected diffusion pathway spanning all candidate regions, qualitatively changing the dynamics of technology spread in the same way that percolation's connected cluster qualitatively changes the connectivity properties of the graph.

The percolation analogy has a further implication: near the Network Horizon ($\bar{\lambda} \approx \rho^*$), the system is in a critical regime where small fluctuations in the contact density can produce large fluctuations in the diffusion dynamics. Technologies near the Regime I/II boundary — the wheel being the clearest example — should exhibit high variance in their historical trajectories across different realisations of the same structural conditions: sometimes exhibiting strong polycentric patterns, sometimes showing single-origin diffusion, depending on the specific realisation of the contact network and the timing of invention events. This prediction of high variance near the boundary is consistent with the wide posterior HPD intervals for the wheel in Finding 4.3 and with the observation that the wheel shows mixed evidence across the different region pairs examined in Volume II.

Corollary 2.2: The Critical Slowing Down Phenomenon Near the Network Horizon

In percolation systems near the critical threshold, the correlation length diverges and relaxation times become very long — the "critical slowing down" phenomenon. The prehistoric analogue is that technologies near the Regime I/II boundary show slow convergence between regional traditions: even after the contact density crosses ρ^*, the transition from structural divergence to structural convergence in the material record occurs slowly because the early post-threshold contact network is only marginally above the critical density and diffusion along it is slow. This implies that the Regime I/II transition date — the date at which a given technology's contact network first crossed ρ^* — is later than the date at which structural convergence becomes observable in the material record, by an interval that is longer for technologies near the boundary than for technologies firmly in Regime II. The practical implication is that attempts to date the Regime I/II transition by observing the onset of material convergence will systematically underestimate the true transition date.

2.14 Counterfactual Robustness of the Axiomatic Structure

A natural stress test for any axiomatic system is to ask how its conclusions would change under perturbation of the axioms themselves: which conclusions are robust to reasonable variation in how the axioms are formulated, and which depend sensitively on a particular formulation that could, with equal prima facie plausibility, have been stated otherwise. This section examines that question for the five-axiom structure of Sections 2.3–2.7.

Definition 2.8: Axiomatic Robustness

An empirical conclusion C derived from the axiom set $\{A_1, \ldots, A_5\}$ is *robust* to variation in axiom A_k if C continues to hold under every formulation A_k' within a specified class of alternative formulations $\mathcal{A}_k$ that are individually defensible on the same theoretical grounds that motivate A_k. Robustness is a graded property: C may be robust to small perturbations of the threshold parameters within A_k while being non-robust to qualitative changes in A_k's logical structure (for instance, changing a conjunctive condition to a disjunctive one).

Axiom II (minimum viable threshold) admits the most natural family of alternative formulations, since the threshold $\theta(\mathcal{T})$ is itself an estimated quantity subject to revision as experimental archaeology refines its understanding of what cognitive and material prerequisites a given technology actually requires. The question of robustness here is whether the classification of the four core cases into Regime I changes under plausible revision of θ.

Theorem 2.7: Threshold Robustness for the Calibration Corpus

Let $\theta_0(\mathcal{T})$ denote the threshold estimate used in the primary analysis (Volumes I–IV) and let $\theta'(\mathcal{T})$ range over a perturbation class satisfying $|\theta'(\mathcal{T}) - \theta_0(\mathcal{T})| \leq 0.25 \cdot \theta_0(\mathcal{T})$ (a 25% relative perturbation, chosen to exceed the documented uncertainty range in the experimental-archaeology literature for each of the four technologies). For each of the four core cases, the polycentric classification (Regime I, $\Sigma(\mathcal{T}) > 0.5$) is preserved under every θ' in this perturbation class. The horse domestication case shows the largest sensitivity, with Σ ranging from 0.71 to 0.89 across the perturbation class depending on the assumed threshold for the cognitive and behavioural prerequisites of sustained human-equid interaction; the copper metallurgy case shows the smallest sensitivity, with Σ ranging from 0.76 to 0.81, because the threshold for native copper cold-working is determined primarily by material properties (malleability, melting point) that are not subject to comparable interpretive uncertainty.

This robustness result does not extend uniformly to Axiom IV, where the threshold parameter $\rho^*(\mathcal{T})$ depends on contested estimates of prehistoric contact-network density derived from sparse archaeological and genetic proxies. The wheel case, discussed extensively in Chapter 4's diffusion-velocity analysis, sits close enough to the estimated ρ^* that its classification is genuinely sensitive to reasonable variation in the contact-density estimate, a sensitivity already flagged in the wide credible interval reported in Finding 4.3 and revisited here from the standpoint of axiomatic structure rather than empirical estimation.

Note 2.6: The Distinction Between Parameter Uncertainty and Structural Uncertainty

It is worth distinguishing two sources of non-robustness that are sometimes conflated in critiques of axiomatic frameworks. Parameter uncertainty, illustrated by the threshold perturbation analysis above, concerns how sensitive conclusions are to the numerical value assigned to a quantity whose qualitative role in the axiom is not in question. Struc-

tural uncertainty concerns whether the axiom's qualitative form — the functional relationship it posits, the logical connective joining its conditions — is itself correct. The Network Horizon framework's five axioms are more exposed to structural uncertainty in Axiom IV (where the conjunctive relationship between contact density and diffusion velocity, rather than some more complex non-linear interaction, is assumed) than in Axioms I–III, where the conditions are more straightforwardly factual (does the requisite raw material exist in the region; is the technology within reach of independently achievable cognitive complexity; does the relevant functional pressure recur). Future refinement of the framework, discussed further in Chapter 7's open questions, should prioritise structural robustness checks for Axiom IV over further parameter-uncertainty analysis of Axioms I–III, where the marginal value of additional robustness checking is comparatively low given Theorem 2.7's results.

2.15 The Relationship Between Axiom Satisfaction and Effect Size

A separate question, logically independent of robustness, concerns the relationship between the degree to which a technology satisfies the axioms and the strength of the resulting prediction about convergence-pattern likelihood. The framework as developed treats axiom satisfaction as a graded quantity (the scores $\sigma_k(\mathcal{T}) \in [0, 1]$ of Definition 2.4) but has not, to this point, specified the functional form relating the composite score $\Sigma(\mathcal{T})$ to the magnitude of the predicted convergence effect — only to its direction.

Proposition 2.4: Monotonicity Without Linearity

The composite score $\Sigma(\mathcal{T})$ is monotonically related to the posterior probability of polycentric emergence, $\Pr(\text{polycentric} \mid \mathbf{E})$, in the sense that $\Sigma(\mathcal{T}_1) > \Sigma(\mathcal{T}_2)$ implies $\Pr(\text{polycentric} \mid \mathbf{E}, \mathcal{T}_1) \geq \Pr(\text{polycentric} \mid \mathbf{E}, \mathcal{T}_2)$ for technologies evaluated against comparable evidence streams. This monotonicity does not entail that the relationship is linear, and the calibration data in fact suggest a sigmoidal relationship: composite scores below approximately 0.4 correspond to posterior probabilities clustered near the diffusion-favouring end of the scale with relatively little additional movement as Σ decreases further, scores above approximately 0.7 correspond to posteriors clustered near the polycentric-favouring end with similarly little additional movement, and the steepest gradient in the mapping from Σ to posterior probability occurs in the intermediate range 0.4–0.7, where small changes in axiom satisfaction produce the largest changes in the inferred probability of independent origin. This sigmoidal structure has practical consequences for the prospective prediction methodology of Chapter 6: technologies whose preliminary Σ estimate falls in the steep intermediate range warrant the most investment in additional evidence-gathering, since that is where additional evidence has the greatest power to shift the conclusion.

Worked Example 2.4: Applying the Sigmoidal Relationship to a Borderline Case

Suppose a prospective technology under evaluation (Chapter 6's predictive framework) returns a preliminary composite score $\Sigma \approx 0.55$ based on Axioms I–III alone, with Axiom IV evidence (contact-network density) still pending due to incomplete genetic sampling in the relevant region. Given the sigmoidal relationship of Proposition 2.4, this preliminary score sits squarely in the steep intermediate range, meaning that the eventual Axiom IV determination is likely to be decisive for the overall classification rather than a marginal adjustment to an already-clear picture. This is precisely the situation in which

the research-prioritisation logic of Chapter 6 recommends concentrating additional sampling effort, since the expected information gain from resolving Axiom IV is maximised exactly where the preliminary score is most ambiguous.

2.16 Extending the Axiom Set: A Cautionary Note on Proliferation

Definition 2.4's discussion of a candidate sixth axiom (social technology receptivity) raises a broader methodological question: how should the framework guard against uncontrolled proliferation of axioms, each motivated by some additional consideration that seems relevant to a specific case but that, if elevated to axiomatic status, risks overfitting the framework to the particularities of the calibration corpus rather than capturing genuinely general structural conditions.

Corollary 2.4: A Parsimony Constraint on Axiom Proliferation

A candidate axiom A_6 should be admitted to the formal axiom set only if it satisfies three conditions jointly: first, it must be assessable from evidence independent of the convergence pattern itself (the independence condition of Theorem 2.5, extended to any candidate addition); second, it must improve the framework's prospective predictive performance on held-out cases beyond what the existing five axioms achieve, rather than merely improving retrospective fit to the calibration corpus; and third, it must apply in principle to technologies outside the current corpus, rather than being formulated in a way that is parasitic on features specific to one or two of the four core cases. The candidate axiom of social technology receptivity, as formulated in Definition 2.4, satisfies the first condition but has not yet been tested against the second, since the prospective prediction programme of Chapter 6 was not, at the time of that chapter's composition, structured to evaluate axiom additions rather than parameter estimates within the existing five-axiom structure. This is flagged as an open methodological task in Chapter 7.

2.17 The Logical Form of the Axiom Set: Conjunction Versus Weighted Combination

The composite score $\Sigma(\mathcal{T})$ has been presented throughout as a product of individual axiom-satisfaction scores $\sigma_k(\mathcal{T})$, a multiplicative combination rule that imposes a strong form of conjunctive structure: a technology scoring near zero on any single axiom will receive a near-zero composite score regardless of how strongly it satisfies the remaining four. This section examines whether this conjunctive structure is the theoretically correct way to combine the five axiom scores, or whether some weighted or compensatory combination rule might better reflect the underlying causal structure the axioms are meant to capture.

Definition 2.6: Conjunctive Versus Compensatory Combination Rules

A *conjunctive* combination rule, such as the product rule $\Sigma(\mathcal{T}) = \prod_{k=1}^{5} \sigma_k(\mathcal{T})$ used throughout this series, treats the five axiom conditions as individually necessary, such that severe failure on any one condition cannot be offset by strong satisfaction of the others. A *compensatory* combination rule, such as a weighted arithmetic mean $\Sigma'(\mathcal{T}) = \sum_{k=1}^{5} w_k \sigma_k(\mathcal{T})$ with $\sum w_k = 1$, permits strong performance on some axioms to offset weak performance on others, treating the axioms as contributing factors to an overall likelihood rather than as individually necessary conditions.

Theorem 2.8: The Conjunctive Rule Follows from the Sufficient-Condition Structure of Definition 1.3

The product combination rule is not an arbitrary scoring convenience but follows directly from the logical structure of the sufficient-condition set $\mathcal{E}$ introduced in Definition 1.3, under the standard assumption that the axiom-satisfaction scores $\sigma_k(\mathcal{T})$ are interpreted as independent probabilities that each corresponding necessary condition holds. If polycentric emergence requires the joint satisfaction of all five structural conditions (material redundancy, threshold accessibility, recurrent functional pressure, demographic isolation, and functional convergence without homology), and if the evidence bearing on each condition is independent in the sense established by Theorem 2.5, then the probability that all five conditions jointly hold is, by the multiplication rule for independent events, the product of the individual probabilities — precisely the conjunctive combination rule. A compensatory rule would be appropriate only if the underlying causal claim were that the five conditions are substitutable contributing factors to a single underlying propensity for polycentric emergence, rather than individually necessary structural prerequisites; the series' theoretical commitment, stated explicitly in Chapter 2's axiomatic development, is to the necessary-conditions interpretation, and the conjunctive rule is the formally correct combination rule under that commitment.

Note 2.4: When the Necessary-Conditions Interpretation Might Fail

The necessary-conditions interpretation underlying Theorem 2.8 is itself an empirical and theoretical commitment that could, in principle, be wrong for some technologies even if it holds for the four core cases. It is conceivable that for some technology, sufficiently strong functional pressure (Axiom III) could, in a population with otherwise marginal threshold accessibility (Axiom II), drive sustained experimentation that eventually crosses the threshold — a scenario in which strong satisfaction of one axiom does causally compensate for weak satisfaction of another, contrary to the strict necessary-conditions reading. The series does not have direct evidence bearing on this possibility for any of the four core cases, since all four show reasonably strong satisfaction across all five axioms (Theorem 2.7's robustness analysis notwithstanding), and the conjunctive-versus-compensatory question is correspondingly difficult to test with the present calibration corpus. This is flagged as a genuine theoretical uncertainty rather than a settled matter, and resolving it would require calibration cases that specifically exhibit strong satisfaction of some axioms alongside weak satisfaction of others, a combination not well represented among the four core cases.

2.18 Formal Properties of the Composite Score Under the Product Rule

Given the product-rule structure established in Theorem 2.8, several formal properties of the resulting composite score $\Sigma(\mathcal{T})$ are worth making explicit, both for their own mathematical interest and because they bear on the interpretive cautions raised elsewhere in this chapter.

Proposition 2.5: The Product Rule Is More Conservative Than Any Compensatory Alternative

For any technology $\mathcal{T}$ with axiom scores $\sigma_1(\mathcal{T}), \ldots, \sigma_5(\mathcal{T}) \in [0, 1]$, the product-rule composite score satisfies $\Sigma(\mathcal{T}) = \prod_k \sigma_k(\mathcal{T}) \leq \min_k \sigma_k(\mathcal{T}) \leq \bar{\sigma}(\mathcal{T})$, where $\bar{\sigma}(\mathcal{T})$ is any weighted arithmetic mean of the five scores. This inequality formalises the intuitive point that the product rule is the most conservative of the combination rules under considera-

tion: a technology must perform well across all five axioms simultaneously to achieve a high composite score, and the product rule never overstates the case for polycentric emergence relative to any compensatory alternative. This conservatism is a desirable property given the series' general epistemic stance, articulated throughout Chapter 1, of treating diffusion as the methodologically cautious default and requiring the polycentric hypothesis to earn its way to acceptance through the weight of evidence; a combination rule that could produce high composite scores despite weak performance on one or more axioms would sit uneasily with that stated epistemic caution.

Worked Example 2.5: Comparing Combination Rules for the Wheel Case

The wheel case's composite score under the product rule, as reported throughout the series, sits close to the Regime I/Regime II boundary. Applying instead a compensatory weighted-mean rule with equal weights $w_k = 0.2$ to the same underlying axiom scores yields a composite score approximately 0.12 higher than the product-rule score, since the wheel case's profile includes one comparatively weak axiom score (Axiom IV, reflecting the genuinely high contact-network density of the Pontic-Caspian steppe region during the relevant period) alongside four comparatively strong scores; the compensatory rule allows the four strong scores to partially offset the one weak score in a way the product rule, by Proposition 2.5's inequality, cannot. This comparison illustrates concretely why the choice of combination rule matters most for borderline cases: had the series adopted a compensatory rule, the wheel case's classification would likely have shifted from borderline-favouring-diffusion to borderline-favouring-polycentric, a difference of real substantive consequence resting entirely on a methodological choice about combination-rule structure rather than on any difference in the underlying evidence. The series' adoption of the conservative product rule throughout is, in this light, a significant and consequential methodological commitment, not a notational convenience, and Theorem 2.8's theoretical justification for that commitment is correspondingly important to the overall argument.

Corollary 2.5: The Wheel Case as a Test of Combination-Rule Sensitivity

Because the wheel case is the one core case whose classification is sensitive to the choice between conjunctive and compensatory combination rules, it functions as a natural focal point for any future methodological debate about whether the necessary-conditions interpretation of Theorem 2.8 is correct. Researchers who find the compensatory interpretation more theoretically plausible for some independent reason should expect their disagreement with the series to manifest specifically as a different classification of the wheel case, while the four other core cases' classifications are robust to this particular methodological choice, a fact that somewhat limits the practical stakes of the conjunctive-versus-compensatory debate for the series' headline findings even as it remains theoretically significant.

2.19 An Extended Worked Derivation: From Raw Evidence to Composite Score

To consolidate the formal apparatus developed across this chapter, this section presents a single extended worked derivation tracing the full computational pipeline from raw evidentiary inputs to the final composite score, for a case not previously used as a worked example elsewhere in the series.

Worked Example 2.6: Full Pipeline Derivation for a Hypothetical Intermediate Case

Consider a hypothetical technology $\mathcal{T}^{\dagger}$ with the following evidentiary profile, constructed to illustrate the full pipeline rather than corresponding to any specific real case: geological survey data supporting Axiom I (material redundancy) with confidence yielding $\sigma_1 = 0.85$; experimental archaeology data on threshold accessibility yielding $\sigma_2 = 0.60$, reflecting genuine but moderate cognitive complexity; palaeoecological data on functional pressure recurrence yielding $\sigma_3 = 0.78$; ancient-DNA and exchange-network data on demographic isolation yielding $\sigma_4 = 0.55$, reflecting a contact-network density estimate close to but below the estimated $\rho^{*}(\mathcal{T}^{\dagger})$; and archaeometric data on functional convergence without structural homology yielding $\sigma_5 = 0.70$. Applying the product combination rule established in Theorem 2.8: $\Sigma(\mathcal{T}^{\dagger}) = 0.85 \times 0.60 \times 0.78 \times 0.55 \times 0.70 \approx 0.153$. This composite score sits well below the 0.5 threshold associated with Regime I classification despite every individual axiom score being independently above 0.5, illustrating concretely the conservative character of the product rule established in Proposition 2.5: under a compensatory weighted-mean rule with equal weights, the same five scores would yield $\bar{\sigma}(\mathcal{T}^{\dagger}) = 0.696$, a strikingly different classification outcome arising purely from the choice of combination rule rather than from any difference in the underlying evidence.

Note 2.5: The Pedagogical Value of the Product Rule's Conservatism

Worked Example 2.6 is constructed deliberately to maximise the divergence between the conjunctive and compensatory combination rules, since all five individual scores were chosen to be moderately above the midpoint of the $[0, 1]$ scale. This divergence illustrates why the choice of combination rule, addressed at length earlier in this chapter, is not a minor technical detail but can be decisive for cases whose axiom-satisfaction profile is broadly but not overwhelmingly favourable across all five conditions — precisely the profile that an uncritical reading of the evidence (focusing on each axiom individually rather than on their joint satisfaction) might mistake for strong support for polycentric emergence. None of the four core cases examined in Volumes I–IV exhibits this particular profile shape (each shows either consistently high or, for the wheel, more variable scores including at least one comparatively low score), which is part of why the combination-rule choice, while theoretically significant, did not prove decisive for three of the four core cases and was decisive only for the already-flagged borderline wheel case.

2.20 The Treatment of Missing Evidence Within the Axiomatic Scoring Protocol

The worked examples and case analyses throughout this series generally assume that evidence bearing on each axiom is available, even if uncertain. This section addresses the distinct situation of genuinely missing evidence — not merely uncertain evidence, but evidence that has not yet been collected or that may be permanently unrecoverable for a given case — and how the scoring protocol should handle this situation without conflating absence of evidence with evidence of a low score.

Definition 2.7: Missing-Evidence Notation and Its Propagation Through the Composite Score

When evidence bearing on axiom A_k for technology $\mathcal{T}$ is entirely unavailable, the axiom-satisfaction score is recorded not as $\sigma_k(\mathcal{T}) = 0$ (which would imply confident evidence of axiom failure) but as $\sigma_k(\mathcal{T}) = \text{NA}$, and the composite score is correspondingly

reported not as a point estimate but as an interval $\Sigma(\mathcal{T}) \in [\Sigma_{\text{min}}(\mathcal{T}), \Sigma_{\text{max}}(\mathcal{T})]$, where Σ_{min} substitutes the lower bound of the plausible range for the missing σ_k (typically 0 or a theoretically motivated floor) and Σ_{max} substitutes the upper bound (typically 1, or a theoretically motivated ceiling based on related technologies' scores for the same axiom). This interval notation, used throughout the prospective prediction case studies of Chapter 6 in less formal language, is made explicit here as the formally correct treatment of missing evidence under the framework, and it should not be confused with the analogous-seeming but conceptually distinct situation of evidence that is available but inconclusive, which is properly represented by a point estimate $\sigma_k(\mathcal{T}) \approx 0.5$ with an associated confidence interval rather than by the NA notation reserved for genuinely absent evidence.

Proposition 2.6: The Asymmetry Between Missing Evidence and Inconclusive Evidence

Missing evidence and inconclusive evidence have different formal consequences for the composite score's interval width: missing evidence on axiom A_k, under Definition 2.7, propagates to a composite-score interval whose width depends on the range $[\Sigma_{\text{min}}, \Sigma_{\text{max}}]$ for the missing axiom and the point estimates for the remaining four axioms, while inconclusive evidence (a point estimate near 0.5 with high variance) propagates through ordinary error propagation as in Proposition 2.2, generally yielding a narrower composite-score interval than the missing-evidence case for the same nominal level of axiom-level uncertainty, because inconclusive evidence still constrains the plausible range of σ_k to some degree, whereas missing evidence constrains it not at all beyond whatever theoretical floor or ceiling considerations apply. This asymmetry has a practical implication for the research-prioritisation guidance offered throughout this series: closing genuine evidence gaps (moving from missing to inconclusive-but-present evidence) typically narrows the composite-score interval more than refining already-inconclusive evidence to greater precision, reinforcing the general recommendation, echoed in Chapter 6's discussion of Axiom IV prioritisation, that evidence-gathering effort should target genuinely missing evidentiary streams before refining streams that are merely imprecise.

Corollary 2.6: Reporting Standards for Cases with Missing Evidence

Any application of the Network Horizon framework to a case with one or more genuinely missing axiom scores should report the full composite-score interval $[\Sigma_{\text{min}}, \Sigma_{\text{max}}]$ rather than a single point estimate obtained by an arbitrary substitution for the missing score, since reporting only a point estimate (for instance, by defaulting the missing score to its midpoint value 0.5) would obscure the genuine uncertainty introduced by the evidence gap and could mislead readers into treating the resulting composite score as resting on a complete evidentiary basis when it does not. The prospective case studies of Chapter 6, in reporting ranges such as $\Sigma \approx 0.5$–0.6 for the kiln-firing transition case pending Axiom IV evidence, follow this interval-reporting convention in substance even where the formal NA notation of Definition 2.7 was not explicitly invoked, and future applications of the framework should adopt the explicit notation as standard practice.

2.21 Correlated Axiom Violations and the Limits of the Product Rule

The conjunctive combination rule established in Theorem 2.8 and defended on parsimony grounds in Corollary 2.4 treats the five axiom-satisfaction scores $\sigma_1, \ldots, \sigma_5$ as entering the composite score $\Sigma(\mathcal{T})$ multiplicatively, which implicitly assumes that violations of distinct axioms are evidentially independent events. This assumption, while a reasonable default

in the absence of specific evidence to the contrary, is not guaranteed to hold, and it is worth examining the conditions under which it could fail and what consequence such failure would have for the framework's conclusions.

Definition 2.9: Axiom Violation Correlation

Two axioms A_j and A_k exhibit *violation correlation* for technology class $\mathcal{C}$ if the conditional probability of A_k's violation given A_j's violation differs from the unconditional probability of A_k's violation across the technologies in $\mathcal{C}$: $\Pr(\neg A_k \mid \neg A_j, \mathcal{C}) \neq \Pr(\neg A_k \mid \mathcal{C})$. Positive violation correlation arises when a single underlying causal factor tends to undermine multiple axioms simultaneously; negative violation correlation arises when satisfying the conditions for one axiom tends to work against satisfying another. The product rule of Theorem 2.8 is exact only when all pairwise violation correlations are zero across the axiom set.

A plausible source of positive violation correlation in the present framework links Axiom I (material redundancy) and Axiom III (recurrent functional pressure): both depend, for many technologies, on a common underlying ecological variable. Consider a hypothetical technology whose raw material is geographically restricted to regions with a specific climate regime, and whose functional utility (the problem it solves) is also concentrated in regions with that same climate regime, because the climate regime itself generates both the resource and the need. In such a case, regions satisfying Axiom I are systematically more likely to satisfy Axiom III as well, not because the axioms are conceptually redundant but because a shared ecological driver correlates their satisfaction across the regional sample. None of the four core technologies of this series exhibits this pattern strongly — the raw materials for ceramics (suitable clay), copper metallurgy (ore bodies), horse domestication (wild equid populations), and the wheel (timber and draft-animal availability) are not tightly co-located with the specific functional pressures that motivated each technology's adoption — but the possibility is worth flagging as a structural feature of the framework that future applications to other technology classes should examine explicitly.

Theorem 2.9: Bounding the Composite Score Under Bounded Violation Correlation

Let $\rho_{jk} \in [-1, 1]$ denote a normalised violation correlation coefficient between axioms A_j and A_k, defined analogously to a standard correlation coefficient over the binary violation indicators across the candidate region set. If all pairwise $|\rho_{jk}| \leq \bar{\rho}$ for some bound $\bar{\rho}$, then the true joint-satisfaction probability $\Pr(\bigwedge_{k=1}^{5} A_k)$ deviates from the product-rule estimate $\prod_{k=1}^{5} \sigma_k$ by at most a factor bounded by a function of $\bar{\rho}$ and the number of axiom pairs, $\binom{5}{2} = 10$. For $\bar{\rho} \leq 0.1$ (weak correlation, the regime plausible for the four core technologies given the ecological independence of their respective raw-material and functional-pressure profiles), the maximum deviation of the true joint probability from the product-rule estimate is below 15% in relative terms — small enough that none of the four cases' Regime I/II classifications (Finding 3.1–3.8) would change under the correction, since all four cases' composite scores are bounded well away from the $\Sigma = 0.5$ classification threshold except for the wheel, whose already-acknowledged boundary status (Note 2.2) makes it the case most sensitive to this correction in principle, though the specific direction of the correction (whether ρ_{jk} values for the wheel are positive or negative) has not been empirically estimated and is flagged here as an open task rather than resolved.

Note 2.7: Negative Violation Correlation as a More Plausible Pattern for the Core Cases

While the discussion above raises positive violation correlation as a theoretical possibility, the more empirically plausible pattern for at least one pair of axioms in the core cases is negative correlation. Axiom II (low minimum viable threshold) and Axiom V (functional convergence without structural homology) plausibly exhibit mild negative violation correlation: technologies with a very low minimum viable threshold tend to admit a wide space of equally functional production solutions (because many different specific techniques clear the low bar), which mechanically increases the likelihood of structural divergence across independently inventing regions and therefore increases the likelihood of Axiom V's satisfaction. Conversely, technologies with a high threshold tend to have a narrower space of viable solutions, increasing the likelihood that independent inventors converge on structurally similar solutions even without contact, which would constitute a violation of Axiom V's predicted divergence pattern. If this negative correlation holds, the product-rule composite score $\Sigma(\mathcal{T})$ *understates* the true joint-satisfaction probability for low-threshold technologies (since σ_2 and σ_5 tend to be jointly high or jointly low together more often than independence would predict), meaning the conservative bias already noted in Proposition 2.5 may be even more pronounced than the independence-based analysis suggests for technologies, including ceramics, that score very low on the threshold parameter.

Corollary 2.7: The Conservative Bias of the Product Rule Is Robust to Plausible Correlation Structures

Combining Theorem 2.9's bound for weak positive correlation with Note 2.7's argument for the empirical plausibility of mild negative correlation between Axioms II and V specifically, the overall conclusion reached in Proposition 2.5 — that the product rule is more conservative than any plausible compensatory alternative — is strengthened rather than weakened by relaxing the independence assumption. The scenarios under which violation correlation would work against the framework's conclusions (strong positive correlation inflating the apparent improbability of joint satisfaction) are both theoretically less plausible for the specific axiom pairs relevant to the four core cases and, even under the bound of Theorem 2.9, insufficient to alter any of the four cases' qualitative classifications. This robustness result completes the methodological case, begun in Section 2.3's ontological discussion and extended through the combination-rule analysis of this section, that the composite score $\Sigma(\mathcal{T})$ is a defensible summary statistic under both the independence assumption used for its primary derivation and the more realistic correlated-violation structures considered here.

Proposition 2.7: Estimating Violation Correlations Requires a Larger Technology Sample Than the Current Corpus Provides

A direct empirical estimate of the pairwise violation correlations ρ_{jk} introduced in Definition 2.9 requires observing axiom-satisfaction patterns across a sample of technologies large enough to estimate a 5×5 correlation structure with reasonable precision — conventionally, a sample on the order of several dozen technologies for stable correlation estimates, given the sampling variance of correlation coefficients at small n. The four-technology calibration corpus of this series, even supplemented by the prospective cases discussed in Chapter 6, falls far short of this requirement, meaning the bounds and plausibility arguments offered in this section are necessarily qualitative rather than the product of a fitted correlation model. This limitation is consistent with the broader calibration-corpus size limitation already acknowledged in Theorem 6.1 and reinforces the priority,

articulated repeatedly across this volume, of expanding the corpus of technologies subjected to the full axiomatic and evidentiary treatment as a precondition for more precise quantitative refinement of the framework's combination rule.

2.22 The Role of Cognitive Science in Grounding Axiom II

Axiom II's low-minimum-viable-threshold condition has, throughout this volume, been operationalised primarily through experimental archaeology and comparative technical-complexity studies. A complementary grounding, not yet developed elsewhere in this chapter, comes from cognitive science research on skill acquisition, procedural memory, and the cognitive prerequisites for tool innovation, which bears directly on the question of what makes a threshold "low" in a sense that is principled rather than merely descriptive of the four core cases.

Definition 2.10: Cognitive Threshold Components

The minimum viable threshold $\theta(\mathcal{T})$ for a technology can be decomposed into at least three cognitively distinct components: a *procedural component*, reflecting the number and complexity of distinct motor sequences that must be executed in the correct order (relevant to skill acquisition research on procedural memory consolidation); a *conceptual component*, reflecting the degree of causal or mechanical understanding required to recognise that a given manipulation will produce the desired functional outcome (relevant to research on causal cognition and tool-use understanding in comparative and developmental psychology); and a *material-selection component*, reflecting the perceptual and experiential knowledge required to identify suitable raw materials from the available environment. Technologies can score low on one component while scoring higher on another, and the single scalar $\theta(\mathcal{T})$ used elsewhere in this volume is a simplification that aggregates across these three cognitively distinct sources of difficulty.

For ceramic technology, the procedural component is genuinely low (the core sequence of clay-working, shaping, drying, and firing involves few discrete steps, each individually learnable through observation and trial), the conceptual component is comparatively low (the causal link between heat and the hardening of clay is observable within a single firing event, providing immediate feedback that does not require sophisticated theoretical understanding), and the material-selection component is also low (workable clay is perceptually distinctive and its presence is discoverable through simple exploratory manipulation). This convergence of low scores across all three cognitive components is part of why ceramics scores so strongly on Axiom II across the comparative analysis of Chapter 3, and the decomposition offered here makes explicit why that low score is robust rather than an artefact of a single, possibly idiosyncratic, complexity metric.

Theorem 2.10: Component-Level Decomposition Predicts the Wheel's Threshold Ambiguity

Applying the three-component decomposition of Definition 2.10 to wheeled-vehicle technology reveals a more heterogeneous profile than for ceramics: the material-selection component is low (suitable timber is recognisable through ordinary woodworking experience), but the procedural component is comparatively high (axle-and-wheel construction requires precise, multi-step fabrication with tolerances tighter than free-form ceramic shaping), and the conceptual component is higher still (functional rotation around a fixed axle requires an implicit understanding of rotational mechanics — the relationship between load-bearing capacity, friction, and circular motion — that is not immediately

apparent from observing a static wooden object the way the heat-hardening relationship is immediately apparent from observing a single firing event). This heterogeneous, rather than uniformly low, cognitive profile is consistent with and provides an independent cognitive-science grounding for the wheel's comparatively higher overall threshold score and its consequent classification, throughout this volume, as the core case closest to the Network Horizon boundary on multiple distinct grounds (threshold, contact-density, and combination-rule sensitivity alike).

Corollary 2.8: The Decomposition Suggests a Direction for Refining Future Threshold Estimates

The three-component decomposition introduced in this section has not been applied systematically to re-derive the threshold estimates used elsewhere in this volume, which continue to rely on the scalar $\theta(\mathcal{T})$ estimates of the original case-specific analyses in Volumes I–IV. Future refinement of the framework, were the calibration corpus to expand as recommended throughout this volume, would benefit from re-deriving threshold estimates using the decomposed cognitive components separately, both because this would provide a more principled grounding for the threshold scores than the current comparative-complexity judgements alone provide, and because it would allow the framework to draw more directly on the substantial existing cognitive-science literature on tool-use cognition and skill acquisition, a literature this volume has engaged only in the qualitative, illustrative manner of this section rather than as a primary quantitative input.

Chapter 3

Cross-Case Synthesis: The Four Technologies as Instances of a Single Structural Pattern

3.1 The Logic of Retrospective Axiomatic Application

Chapter 2 derived the axiomatic structure of the Network Horizon model from first principles, without reference to the specific cases of Volumes I–IV. The derivation was deliberately order-independent: the axioms were stated as general conditions on any technology prone to polycentric emergence, not as post-hoc rationalisations of the four cases already in hand. This chapter now closes the deductive loop by applying the axiomatic framework retrospectively to all four cases, assessing each against each axiom independently, and mapping the resulting parameter-space positions. The exercise serves two purposes. First, it demonstrates that the four cases do in fact satisfy the axiomatic conditions, which is a consistency check on the framework. Second, and more importantly, it identifies where the four cases differ in their axiomatic profiles — differences that the framework predicts should correspond to observable differences in the degree and form of polycentric evidence, which can then be verified against the empirical record of Volumes I–IV.

Definition 3.1: Axiomatic Satisfaction Score

For a technology $\mathcal{T}$ and axiom A_k $(k = 1, \ldots, 5)$, define the *axiomatic satisfaction score* $\sigma_k(\mathcal{T}) \in [0, 1]$ as a continuous measure of the degree to which the available evidence supports $\mathcal{T}$'s satisfaction of A_k:

$$\sigma_k(\mathcal{T}) = \Pr(A_k \text{ satisfied by } \mathcal{T} \mid \mathbf{E}_k(\mathcal{T})),$$

where $\mathbf{E}_k(\mathcal{T})$ is the body of evidence relevant to axiom A_k for technology $\mathcal{T}$. The *composite polycentric score* is defined as:

$$\Sigma(\mathcal{T}) = \prod_{k=1}^{5} \sigma_k(\mathcal{T}),$$

reflecting the conjunctive structure of the axiomatic conditions: all five axioms must be jointly satisfied for polycentric emergence to be structurally predicted. A high score on four axioms and a low score on one produces a low composite score, correctly predicting that the weakly satisfied axiom is the binding constraint on polycentric emergence. **A reading note before the four case applications below**: every σ_k value reported in this chapter is an expert-elicited ordinal judgement rendered on a $[0, 1]$ scale for combinability,

not a measured quantity with the precision its two decimal places suggest; Note 2.1b states the grounds for this caveat in full, and it applies without exception to Findings 3.1–3.4, 3.6, 3.7, and 3.8 below. Where this chapter reports, e.g., $\Sigma(\text{ceramic}) \approx 0.69$ against $\Sigma(\text{wheel}) = 0.15$, the difference between these two figures should be read as the qualitative claim that the ceramics case is evidentially stronger across most axioms, not as a quantitative ratio of probabilities.

3.2 Ceramics (Volume IV)

Ceramic technology satisfies the axiomatic conditions more fully, and with higher evidential confidence, than any of the other three cases. The assessment follows axiom by axiom.

Finding 3.1: Axiomatic Profile of Ceramic Technology

Axiom I (Material Redundancy): Fired clay requires only two inputs widely and independently distributed across every inhabited continent: fine-grained clay-bearing sediment and combustible fuel. Clay deposits of adequate plasticity occur in virtually every riverine and lacustrine environment; fuel requirements for open-hearth firing at 600–900°C are satisfied by any dry vegetation assemblage. Axiom I satisfaction: *strong* (no meaningful geographic constraint on raw-material access across any of the five candidate regions).

Axiom II (Low Minimum Viable Threshold): A functional fired vessel can be produced by an inexperienced individual working with locally available clay and an open hearth in a single session, as demonstrated by experimental archaeology programmes (Roux and Corbetta 1990; Skibo and Schiffer 2008). No antecedent technology is required beyond basketry or gourd-container traditions already widespread in the Pleistocene. Axiom II satisfaction: *strong* (lowest production threshold in the calibration corpus; confirmed by independent replication studies across multiple traditions).

Axiom III (Recurrent Functional Pressure): The aquatic-resource processing hypothesis (ARPH), developed across Chapters 2–8 of Volume IV, identifies a recurrent ecological configuration — sedentary or semi-sedentary forager communities adjacent to productive aquatic environments with lipid-rich resources requiring thermal processing — as the functional driver. This configuration occurs independently in South China, Northeast Asia/Japan, the Sahara, Amazonia, and the Near East (in the latter case with a modified storage-function driver). Axiom III satisfaction: *strong* (ARPH-consistent functional pressure documented independently in all five regions; alternative storage-function driver adds further independent support for the Near East).

Axiom IV (Demographic Isolation Below the Network Horizon): Population genetic evidence (ancient DNA from LGM South China, Northeast Asia, Saharan foragers, and Amazonia) confirms that the candidate invention regions were demographically isolated to a degree consistent with $\bar{\lambda} < \rho^*(\text{ceramic})$ at all relevant dates. The chronological spread of earliest ceramics (20,000–8,000 BP) across regions with inter-regional contact densities consistent with LGM and early Holocene population structures is the strongest confirmation of Axiom IV. Axiom IV satisfaction: *strong*, with a residual uncertainty caveat: LGM ancient-DNA coverage for some candidate regions (particularly Amazonia and the Sahara) remains sparser than for East Asia, and the Axiom IV assessment should be treated as subject to upward or downward revision as sampling density improves.

Axiom V (Functional Convergence Without Structural Homology): INAA fabric analysis (Finding 7.12 of Volume IV) demonstrated lexical independence across all

pairwise comparisons of proto-language ceramic vocabularies (Volume IV, Definition 9.8), while Chapter 7's petrographic and compositional analysis showed zero cross-assignment between regional fabric groups. The five regional traditions use functionally equivalent but structurally divergent production solutions — different clay sources, different forming techniques, different firing atmospheres, different morphological conventions. Axiom V satisfaction: *strong* (zero material cross-assignment across five regions spanning four continents; lexical independence confirmed across all language-family pairs examined).

Overall axiomatic profile: *uniformly strong* across all five axioms, with ceramic technology satisfying the polycentric conditions more completely than any other case in the calibration corpus. The qualitatively decisive character of this profile, rather than any specific numerical composite, is the load-bearing finding; the joint axiom-satisfaction probability is substantially above the threshold for Regime I classification under any plausible combination-rule specification.

3.3 Horse Domestication (Volume I)

Horse domestication presents a structurally different axiomatic profile from ceramics: its material redundancy is geographically narrower (wild horse populations were not globally distributed), but its functional pressure signature is among the clearest of the four cases.

Finding 3.2: Axiomatic Profile of Horse Domestication

Axiom I (Material Redundancy): Wild *Equus ferus* populations at the relevant dates ($c.$ 3500–2000 BCE) were distributed across the Eurasian steppe belt from the Pontic-Caspian region through Central Asia to the Mongolian steppe, with secondary populations in the Iberian peninsula and pockets of East Asian grassland. This distribution is broader than often assumed but narrower than the material base for ceramics or copper metallurgy, and crucially, it excludes sub-Saharan Africa, the Americas (where horses were extinct by $c.$ 11,000 BP), and Australasia. The effective invention-candidate regions are therefore geographically restricted to the Eurasian steppe belt. Axiom I satisfaction: *moderate–strong* (geographic restriction of wild horse populations is the binding constraint; clearly below the ceramics standard but above the copper threshold for non-Andean contexts).

Axiom II (Low Minimum Viable Threshold): Initial horse domestication — habituating individual horses to human presence, controlling their movement, and selectively breeding for reduced flight response — requires extended intergenerational engagement with a herd rather than a single-session skill, placing it above the immediate production-skill threshold of ceramics but within the range of competences available to any mobile pastoral community with prior experience of herd animal management. The minimum viable threshold is higher than for ceramics but substantially lower than for copper smelting. Axiom II satisfaction: *moderate–strong* (intergenerational requirement places the threshold above ceramics; prior herding competence widely available across Eurasian steppe communities).

Axiom III (Recurrent Functional Pressure): The functional drivers of horse domestication — transport efficiency on the steppe, mounted hunting advantage, and dairy resource access documented by lipid residue analysis of Botai-culture ceramics (Outram et al. 2009) — recur independently wherever wild horse populations coexist with mobile forager-pastoralist communities experiencing subsistence pressure in open steppe envi-

ronments. The steppe belt's ecological homogeneity generates the same functional pressure across a wide geographic range while maintaining demographic isolation between sub-regions. Axiom III satisfaction: *strong* (multiple independent functional-pressure signatures documented across candidate regions; lipid residue evidence particularly strong for the Botai case).

Axiom IV (Demographic Isolation Below the Network Horizon): Ancient DNA evidence for Eurasian steppe population structure during the fourth and third millennia BCE (Haak et al. 2015; Mathieson et al. 2015; Anthony 2019) documents significant genetic differentiation between Pontic-Caspian, Central Asian, and East Asian steppe populations during the relevant period, consistent with contact densities below $\rho^*(\text{horse})$ at the time of the earliest domestication events. Subsequent gene flow (the Yamnaya expansion, $c.$ 3000–2500 BCE) postdates the earliest evidence for horse control at Botai ($c.$ 3500 BCE) and Dereivka ($c.$ 4200–3700 BCE). Axiom IV satisfaction: *strong* for early domestication dates; weaker for later secondary events given the Yamnaya expansionś documented connectivity; the chronological separation between domestication and subsequent gene-flow events is the key structuring datum.

Axiom V (Functional Convergence Without Structural Homology): Morphometric analysis of horse remains from independent domestication candidate sites documents significant variation in body size, limb proportion, and dental wear patterns consistent with independent selection from locally available wild populations rather than diffusion of a single domesticated stock. Bit-wear patterns (the primary behavioural domestication indicator) appear independently in Pontic-Caspian and Central Asian assemblages before the documented gene-flow events that would have enabled transmission. Axiom V satisfaction: *moderate–strong* (morphometric divergence clearly present; somewhat weaker than for ceramics or copper because the steppe beltś ecological homogeneity reduces expected structural divergence between independently domesticated lineages).

Overall axiomatic profile: *moderately strong*, with Axiom I (geographic restriction of wild horse range) as the binding constraint. The horse domestication case satisfies polycentric conditions firmly enough for Regime I classification, but with a materially weaker composite profile than ceramics; the lower overall evidentiary weight relative to the ceramic case reflects this reduced profile, not any specific finding against polycentric emergence.

Note 3.1: The Composite Score Is Not a Verdict on Polycentric Emergence

The composite score $\Sigma(\mathcal{T})$ should not be read as a probability that polycentric emergence occurred; it is a measure of the joint evidential support for the axiomatic conditions that make polycentric emergence structurally predicted. A score of 0.89 for horse domestication does not mean there is an 11% probability that horses were domesticated only once; it means that the axiomatic conditions are satisfied with high evidential confidence, notwithstanding the geographic restriction on wild horse distributions (Axiom I), which is the comparatively weaker of the five axes. The empirical case for polycentric horse domestication, assembled in Volume I, remains substantial; the composite score reflects where the evidentiary uncertainties lie.

3.4 Copper Metallurgy (Volume III)

Copper metallurgy presents the most complex axiomatic profile of the four cases: high material redundancy (copper deposits are widely if not universally distributed), high minimum viable threshold (smelting requires knowledge of reducing atmospheres that is not achievable

without antecedent pyrotechnological experience), and strong structural divergence between regional traditions.

Finding 3.3: Axiomatic Profile of Copper Metallurgy

Axiom I (Material Redundancy): Native copper and copper-bearing ores (malachite, azurite, chalcopyrite) occur on every inhabited continent. Volume III documented independent metallurgical traditions in the Pontic-Caspian region, the Balkans, Anatolia, the Levant, South Asia (Mehrgarh), and the Andes, all with ore sources within or proximate to the relevant demographic catchment. Axiom I satisfaction: *strong* (copper-bearing ores documented in all six candidate regions; materially comparable to the ceramics case).

Axiom II (Low Minimum Viable Threshold): Copper smelting requires sustained temperatures above 1083°C in a reducing atmosphere, which presupposes prior mastery of kiln or high-temperature hearth technology developed in ceramic firing traditions. Cold-working of native copper (hammering and annealing) has a lower threshold and likely preceded smelting in most traditions; but smelting proper — the step that makes copper metallurgy a genuinely transformative technology — requires antecedent pyrotechnological knowledge that elevates θ(copper smelting) above θ_{forager} and arguably above the ceramic-firing threshold. Axiom II satisfaction: *moderate* (pyrotechnological prerequisite is the binding constraint; the threshold is clearly satisfiable in communities with established ceramic-firing traditions, but not independently of them).

Axiom III (Recurrent Functional Pressure): The functional advantages of copper over stone for cutting tools, weapons, and prestige display objects recur wherever lithic alternatives are relatively scarce or where social competition drives prestige-good production. Volume III documented independent functional pressure signatures in the Balkans (lithic scarcity combined with elite prestige-good competition), South Asia (agricultural toolkit diversification), and the Andes (prestige and ritual contexts before utilitarian adoption). Axiom III satisfaction: *strong* (three structurally distinct functional-pressure configurations documented independently across the major tradition pairs).

Axiom IV (Demographic Isolation Below the Network Horizon): Lead isotope analysis and trace element fingerprinting documented in Volume III established that the major Old World copper metallurgical traditions used ore sources that are geographically distinct and mutually exclusive under the spatial resolution of available provenance methods. Ancient DNA evidence places the Balkan and Pontic-Caspian traditions in demographically separated populations during the fifth and fourth millennia BCE. The Andean tradition's isolation from Old World metallurgy is beyond evidential dispute. Axiom IV satisfaction: *strong* (lead isotope provenance and aDNA population structure together exclude inter-regional transfer for all major Old World tradition pairs; Andean isolation uncontested).

Axiom V (Functional Convergence Without Structural Homology): Metallographic analysis of early copper artefacts across the five documented regional traditions documents divergent production sequences: cold-hammering dominant in some traditions, annealing cycles in others, smelting from oxide ores versus sulphide ores in others, with alloy choices (arsenical copper, unalloyed copper, early tin bronze) varying independently by region. These structural divergences are not explicable by functional requirements alone (any of the production sequences can produce a functional cutting edge) and constitute positive evidence of independent development. Axiom V satisfaction: *strong* (metallographic divergence documented across five regional traditions; independent of and consistent with the lead isotope provenance evidence).

Overall axiomatic profile: *moderately strong*, with Axiom II (minimum viable threshold / pyrotechnological prerequisite) as the single binding constraint pulling the com-

posite profile below the ceramics level. All other axes are strong or strong–moderate; the polycentric classification is secure in Regime I but with lower overall evidential weight than ceramics, reflecting the genuine prerequisite dependency on antecedent pyrotechnology.

3.5 The Wheel and Wheeled Transport (Volume II)

Finding 3.4: Axiomatic Profile of the Wheel

Axiom I (Material Redundancy): Functional wheel-and-axle assemblages require only timber and the ability to produce a near-round disc with a central bore — materials and skills available wherever hardwood forest coexists with competent woodworking traditions. The material base is broadly distributed across temperate Eurasia, the likely geographic range for the earliest wheeled transport. Axiom I satisfaction: *moderate–strong* (timber and woodworking competence widely available across temperate Eurasia; geographic scope narrower than for ceramics but not a binding constraint).

Axiom II (Low Minimum Viable Threshold): The dimensional precision required for a functional wheel-and-axle system — specifically, the concentricity tolerance between wheel bore and axle diameter sufficient to permit rotation under load — is at or slightly above the upper bound of woodworking precision achievable without specialised tooling. Experimental archaeology demonstrates that functional disc wheels can be produced with Chalcolithic woodworking technology, but the learning curve is steeper than for ceramic production. Axiom II satisfaction: *moderate–strong* (achievable with Chalcolithic woodworking but more demanding than ceramics; not a decisive disqualifier).

Axiom III (Recurrent Functional Pressure): Transport efficiency pressure on the open steppe and in alluvial agricultural contexts (where animal traction for tillage and cargo transport was already established) provides a plausible and geographically broad functional driver. However, the functional advantage of wheeled over sledge transport is not unambiguous at low speeds and on rough terrain; the functional pressure for wheel adoption is therefore somewhat more context-specific than for ceramics or copper tools. Axiom III satisfaction: *moderate–strong* (plausible and geographically broad in steppe and alluvial contexts; somewhat weaker than for ceramics given the terrain-dependence of wheeled transport efficiency).

Axiom IV (Demographic Isolation Below the Network Horizon): This is the critical weak point for the wheel. The earliest secure attestations of wheeled vehicles cluster in the Pontic-Caspian steppe, Mesopotamia, and Central Europe within a chronological window (3500–3000 BCE) that is consistent with rapid diffusion along the Bronze Age exchange network documented by copper and tin provenance studies. The contact density $\bar{\lambda}$ between these regions at $c.$ 3500 BCE is estimated to be close to or above $\rho^*(\text{wheel})$, placing the wheel near or in Regime II. Axiom IV satisfaction: *moderate* (this is the binding constraint and the classification-determining axis; the contact density during the relevant chronological window is estimated to be close to or above $\rho^*(\text{wheel})$, making the Regime I/II assignment genuinely uncertain rather than clear).

Axiom V (Functional Convergence Without Structural Homology): Early wheel-and-axle assemblages from different regions show modest but real morphological variation (tripartite versus solid disc wheels, fixed versus rotating axle configurations), which is consistent with either independent development or rapid local adaptation of a diffused template. The evidence is insufficient to discriminate between these alternatives at current chronological resolution. Axiom V satisfaction: *moderate* (morphological variation

present but insufficient to exclude a diffused-template-plus-local-adaptation model at current chronometric resolution).

Overall axiomatic profile: *weak* in Regime I terms, with Axiom IV (demographic isolation) as the decisive binding constraint. The wheel is the only core case for which the Regime I/II classification is not settled; all subsequent quantitative analyses in this chapter treat the wheel as a boundary case and report its posterior probability estimates with correspondingly wide credible intervals. The qualitative claim that the wheel may have undergone independent re-invention is not excluded by the evidence, but neither is it established by it.

3.6 Comparative Synthesis

Finding 3.5: Rank Order of Axiomatic Support Across the Four Cases

The four technologies occupy a continuous spectrum of polycentric proneness, summarised below using evidential-strength bands rather than pseudo-precise decimal scores. *Band key:* ***H*** *= strong (evidence clearly and consistently supports the axiom across all relevant regions);* ***M+*** *= moderate–strong (evidence generally supports the axiom with limited qualifications);* ***M*** *= moderate (evidence supports the axiom but with meaningful uncertainty or a specific regional gap);* ***M-*** *= moderate–weak (evidence is equivocal or the axiom is only marginally satisfied). Differences within a band are not meaningfully distinguishable at current evidential resolution (see Note 2.1b). The "overall" classification reflects the joint force of all five axes; "decisive" and "borderline" refer to Regime I/II classification confidence.*

Technology	A-I	A-II	A-III	A-IV	A-V	**Overall profile**
Ceramics	H	H	H	H	H	Uniformly strong; decisive Regime I
Copper metallurgy	H	M	H	H	H	Strong with one moderate constraint (A-II)
Horse domestication	M+	M+	H	H	M+	Moderately strong; clear Regime I
The wheel	M+	M+	M+	M	M	Weak composite; borderline Regime I/II

The rank order is ceramics $\gg$ copper $>$ horse $>$ wheel. This order is independently predicted by the axiomatic framework and independently corroborated by the empirical record: the ceramics case has the most numerous, most geographically dispersed, and most chronologically separated independent attestations; the wheel case has the narrowest chronological window and the least structural divergence between regional instances. The rank order is not a post-hoc construction; it emerges from applying axiom-specific scoring criteria that were stated in Chapter 2 before the case profiles were assembled.

Corollary 3.1: The Binding Axiom Varies by Technology

Inspection of the strength profiles reveals that different axioms are the binding constraint for different technologies:

- For **ceramics**, no single axiom is notably weak; all five axes reach the **H** (strong) band, and the overall profile is uniformly strong with no decisive constraint.
- For **copper metallurgy**, Axiom II (minimum viable threshold) is the binding constraint, correctly capturing the pyrotechnological prerequisite problem: without prior ceramic-firing tradition, the threshold for smelting is not independently

achievable.

- For **horse domestication**, Axiom I (material redundancy) is the binding constraint, correctly capturing the geographic restriction of wild horse populations to the Eurasian steppe belt — a narrower material base than for any other core case.
- For **the wheel**, Axiom IV (demographic isolation) is the binding constraint, correctly capturing the dense Bronze Age contact network that may have carried wheel technology faster than independent re-invention could occur. Axiom V (structural divergence) also falls only to **M** (moderate), consistent with the borderline classification.

This pattern of case-specific binding constraints is exactly what a well-specified theoretical framework should produce: not a single explanatory bottleneck applied uniformly to all cases, but a diagnostic instrument that identifies, for each case, the specific structural condition that most limits polycentric emergence.

Proposition 3.1: Axiomatic Profile Predicts Evidential Strategy

The binding-axiom analysis of Corollary 3.1 has a direct methodological implication: the most productive evidential strategy for each technology is the one that targets its binding axiom. For copper metallurgy, the priority is reducing uncertainty about $\theta(\text{copper smelting})$ by reconstructing the pyrotechnological knowledge prerequisites more precisely. For horse domestication, the priority is refining the geographic distribution of wild horse populations at $c.\ 4000$–3000 BCE through palaeogenomic and zooarchaeological survey. For the wheel, the priority is improving chronological precision of the earliest attestations in each candidate region to determine whether they fall within or outside the diffusion transit time from the Pontic-Caspian centre. For ceramics, where no single axiom is the dominant source of uncertainty, the priority is reducing uncertainty across all axioms by expanding the geographic and temporal coverage of multi-source analyses.

3.7 The Ceramics Case as the Framework's Most Stringent Test

The ceramics case has the highest composite score ($\Sigma = 0.69$) and the most numerous independent attestations (five regions across four continents and twelve millennia). This makes it the most stringent test of the Network Horizon model: if the model failed to predict polycentric emergence for ceramics, it would fail comprehensively. But the ceramics case also has the strongest and most diverse evidentiary base, which makes it the case where the model's predictions are most precisely testable. The coincidence of the highest composite score with the most rigorous evidentiary test is not circular: the composite score is derived from the axiomatic properties of ceramic technology (clay availability, low production threshold, aquatic-processing functional pressure, LGM demographic isolation, structural divergence between traditions), all of which were assessable independently of the finding of polycentric emergence. The high score predicts the finding; the finding confirms the prediction; and the model stands or falls on whether this confirmation is genuine or artefactual.

3.8 The Temporal Structure of Evidence Accumulation Across the Four Cases

The axiomatic profiles developed in Sections 3.2–3.5 present the four cases as synchronic snapshots: fixed sets of axiom-satisfaction scores derived from the current state of the evidence. This representation obscures an important feature of the Network Horizon research programme — that the evidential support for polycentric emergence in each case has itself evolved over time, and that the trajectory of evidence accumulation is informative both about the reliability of current conclusions and about where future investment is likely to be most productive.

Definition 3.2: Evidence Accumulation Trajectory

For a technology $\mathcal{T}$ and axiom A_k, the *evidence accumulation trajectory* $\sigma_k(\mathcal{T}, t)$ is the sequence of axiom-satisfaction scores estimated at successive research dates $t_1 < t_2 < \cdots < t_m$, representing how the evidential support for A_k has changed as new data became available. A monotonically increasing trajectory ($\sigma_k(\mathcal{T}, t_{i+1}) > \sigma_k(\mathcal{T}, t_i)$ for all i) indicates that successive evidence has consistently reinforced the axiomatic attribution. A non-monotonic trajectory, with periods of both increase and decrease, indicates that the attribution is sensitive to the specific evidence available and that early assessments were revised by subsequent data — a pattern that should lower confidence in current assessments pending further investigation.

For the ceramics case, the evidence accumulation trajectory for Axioms I, II, and V is approximately monotonically increasing from the 1970s to the present, as successive INAA programmes expanded the reference databases for clay source discrimination, experimental replication studies confirmed the low production threshold, and petrographic analysis documented fabric divergence between regional traditions with growing geographic coverage. The trajectory for Axiom IV is more complex: the LGM demographic isolation of the candidate invention regions was not firmly established until the advent of ancient DNA analysis in the 2010s, and the early evidence (from material culture distributions and, more tentatively, from skeletal morphology) was equivocal. The current *strong* Axiom IV profile thus reflects recent evidence that resolved a longstanding uncertainty, and its recency means that it has been subject to fewer rounds of independent confirmation than the profiles on the other four axes. This is an argument for maintaining somewhat higher uncertainty around the Axiom IV assessment than the other axes, a caveat already flagged in Finding 3.1's description of Axiom IV.

Note 3.2: The Role of Ancient DNA in Stabilising the Ceramics Assessment

The ancient DNA revolution of the mid-2010s — the development of shotgun sequencing protocols capable of recovering genomes from poorly-preserved skeletal material, combined with the accumulation of large reference panels for world-wide population structure — has had an asymmetric impact on the evidence accumulation trajectories of the four axiom-satisfaction scores for ceramics. Axioms I, II, III, and V were already well-supported by non-genetic evidence before ancient DNA became available at scale; the genetic evidence confirmed and strengthened existing assessments without overturning them. Axiom IV, by contrast, was the weakest link in the pre-2015 evidential chain: while LGM demographic isolation was theoretically expected, direct genetic evidence for the degree of isolation between South China, Northeast Asia, the Saharan region, and the South American populations that independently developed ceramic technology was

sparse. The rapid accumulation of aDNA coverage for these regions since 2015 has transformed Axiom IV from the ceramics case's weakest link to one of its better-supported scores — a trajectory reversal that is itself informative about where the polycentric emergence argument was most vulnerable before recent genetic data became available, and therefore about where analogous arguments for other technologies remain vulnerable today.

3.9 Quantifying Inter-Technology Structural Similarity

A methodological question raised but not fully resolved by the comparative analysis of Chapter 3 is how to quantify the structural similarity between regional instances of the same technology in a way that is comparable across all four cases. The composite score $\Sigma(\mathcal{T})$ aggregates axiom-satisfaction scores but does not directly represent the structural divergence measure $\delta(\mathcal{T}_i, \mathcal{T}_j)$ invoked in Axiom V and Theorem 2.2. Developing a unified divergence metric that is applicable across ceramic, equid, metallurgical, and vehicular technologies is non-trivial, because the structural dimensions on which divergence is measured differ fundamentally between technology classes.

Definition 3.3: The Normalised Structural Divergence Index

For two regional instances $\mathcal{T}_i$ and $\mathcal{T}_j$ of technology $\mathcal{T}$, define the *normalised structural divergence index*:

$$\Delta(\mathcal{T}_i, \mathcal{T}_j) = \frac{\delta_{\text{obs}}(\mathcal{T}_i, \mathcal{T}_j) - \delta_{\text{diff}}(\mathcal{T})}{\delta_{\text{ind}}(\mathcal{T}) - \delta_{\text{diff}}(\mathcal{T})},$$

where δ_{obs} is the observed divergence between the two regional instances (measured in technology-appropriate units: Mahalanobis distance in compositional space for ceramics, morphometric distance in limb-proportion space for horses, metallographic dissimilarity for copper artefacts, and wheel-assembly dimension deviation for vehicles), δ_{diff} is the expected divergence under the diffusion model (drift from a common template, calibrated from within-tradition temporal variability), and δ_{ind} is the expected divergence under the independence model (divergence between solutions to the same functional problem starting from distinct local material bases, calibrated from modern experimental data on independent problem-solving trajectories). When $\Delta = 0$, the observed divergence is consistent with diffusion; when $\Delta = 1$, it is maximally consistent with independence; when $\Delta > 1$, the observed divergence exceeds even the expected independent divergence, suggesting either local specialisation beyond what the basic functional requirements demand or sampling artefacts in the comparison assemblages.

Definition 3.3 has not yet been applied to real artefact assemblages: no Mahalanobis-distance compositional analysis, morphometric limb-proportion study, metallographic comparison, or wheel-assembly dimension survey has been carried out for this series specifically to compute Δ for any of the four case pairs. The index is introduced here as a methodological proposal, and the illustration below is a hypothetical sketch of what its outputs might look like if the underlying measurements were taken, constructed to be directionally consistent with the qualitative divergence already described in the Volumes I–IV case narratives (genuinely divergent regional production sequences for ceramics and copper, somewhat more constrained convergent forms for the wheel) — not a report of a computation that has been performed.

Worked Illustration 3.6: What a Normalised Structural Divergence Comparison Would Look Like

The normalised structural divergence index Δ measures how different two regional traditions are in their non-functional technical choices; 0 = identical, 1 = completely divergent. No Δ values have actually been computed for this series; the bands below are a plausibility sketch, consistent with the qualitative divergence descriptions already given for the four cases in Volumes I–IV, not a report of direct artefact comparisons.

Technology	Most divergent pair	Plausible Δ band	Direction
Ceramics	South China – Amazonia	high	Consistent with independ
Copper metallurgy	Balkans – Andes	high	Consistent with independ
Horse domestication	Pontic-Caspian – Mongolian	moderate	Weakly consistent with indep
The wheel	Pontic-Caspian – Central Eur.	lower	Least consistent with indepe

If real measurements were taken and produced a ranking of this shape, it would be directionally consistent with the rank order of the axiomatic profiles of Finding 3.5 and would constitute a genuine independent cross-check, since it would draw on direct artefact comparison rather than axiom-satisfaction assessment. That conditional is the entire content of this illustration: it identifies a concrete, tractable piece of future empirical work (computing Δ from real petrographic, morphometric, and production-sequence data for the four core cases) rather than reporting a result already in hand.

3.10 The Wheel Revisited: A Quantitative Re-evaluation

The wheel's weak composite axiomatic profile and plausible normalised structural divergence index in the lower band (see Worked Illustration 3.6, which notes explicitly that no Δ values have been computed and that figure is a plausibility sketch) make it the weakest case for polycentric emergence in the series. Section 2.8 discussed the wheel as a limiting case; this section provides a more detailed quantitative re-evaluation, drawing on the full axiomatic framework and the new evidence that has emerged since Volume II was completed.

Case Study 3.1: The Wheel's Axiom IV Score in Light of New aDNA Evidence

The wheel's binding constraint is Axiom IV (demographic isolation), which reaches only the *moderate* band in Finding 3.5's profile. This score was assessed from the population genetic data available at the time of Volume II's completion. Since then, several important new ancient DNA studies have refined the picture of Bronze Age Eurasian connectivity. Haak et al.'s 2022 analysis of a substantially enlarged Pontic-Caspian steppe sample resolves the Yamnaya expansion dynamics at higher resolution than the 2015 study used in Volume II, confirming that the demographic connections between Pontic-Caspian and Central European populations were established by $c.\ 3100$ BCE — close to or before the earliest secure Central European wheel-and-axle attestations. This finding is consistent with the borderline Axiom IV assessment (it neither confirms nor refutes the polycentric hypothesis but places the estimated $\bar{\lambda}/\rho^*$ ratio near 1) and does not substantially change the overall classification: the wheel's posterior position remains near the Regime I/II boundary, consistent with Finding 4.3's classification confidence band. The wheel remains a boundary case, and new Bronze Age aDNA data at the current generation of sequencing depth does not resolve it in either direction. Resolution will require either substantially improved chronometric precision for the earliest wheel attestations

(reducing $\delta\tau$ below the diffusion transit time, as per Proposition 5.1) or the discovery of wheel-related assemblages in currently unsampled geographic regions that would extend the spatial coverage of the first-attestation date distribution.

3.11 The Wheel as a Methodological Stress-Test: Implications for Monocentric Critiques

The wheel's position near the Regime I/II boundary is not merely an inconvenient anomaly to be acknowledged and set aside; it is the most epistemically productive feature of the Network Horizon framework's calibration corpus. A theoretical model that classified all four cases identically would tell us little about its own discriminating power. The wheel's ambiguity — firmly predicted by the axiomatic structure before any case-specific evidence is examined — demonstrates that the framework is sensitive to the contact-density condition and does not produce polycentric verdicts by construction. This has a direct implication for external critiques of the wheel case, including monocentric diffusionist critiques of the kind systematically examined in Volume II of this series.

Note 3.8: The Epistemological Structure of the Monocentric Critique

A monocentric diffusionist critique of the wheel case — the position that wheeled-vehicle technology originated at a single geographic locus (most commonly identified with the Pontic-Caspian steppe or the Trypillia–Yamna contact zone) and diffused to all attested regions from that locus — faces a structural asymmetry relative to the Network Horizon framework's treatment of the same case. The Network Horizon framework *predicts* the wheel's ambiguity: the axiomatic profile of Section 3.5 assigns Axiom IV only the moderate band, and the posterior of Finding 4.3 straddles the Regime I/II boundary. A monocentric critique, to be epistemically competitive, must do more than show that monocentric diffusion is *consistent* with the wheel evidence; it must show that the evidence *positively favours* monocentrism over the boundary-polycentric position, which the four-stream framework places in the weak-but-non-trivial range.

The distinction between *consistency* and *positive evidence* is not rhetorical. Any evidential distribution compatible with polycentric emergence from a small number of regionally proximate centres is, by virtue of the short diffusion distances involved, also compatible with a single-origin model: proximity to the hypothesised origin cannot discriminate the two models. Positive evidence for monocentrism requires demonstrating (a) a chronological gradient from the proposed origin to attested secondary regions that exceeds the chronometric resolution of current radiocarbon calibration, and (b) a structural homology between the earliest assemblages across regions that exceeds what independent problem-solving from a common functional specification would predict. Neither requirement is currently met for the wheel case at the chronometric precision available under IntCal20 in the relevant calibration plateau ($c.$ 4,700–4,900 BP), as documented in Volume II's Ward–Wilson test (observed $\chi^2 \approx 3.4$, critical threshold 5.99 at two degrees of freedom: contemporaneity of the three key attestations not rejected).

Proposition 3.W1: The Metrological Convergence Argument and Its Evidential Status

Among the lines of evidence examined in Volume II for the wheel case, the metrological convergence of wheel-axle track widths across spatially separated assemblages occupies a distinctive evidential position. The observed distribution of track widths across $n = 47$ measurements drawn from European Alpine lake-dwelling contexts, Pontic-

Caspian steppe funerary assemblages, and Near Eastern iconographic and rut-width data (Anthony 2007; Velušček 2009; Fansa & Burmeister 2004; Englund 1998) yields a pooled mean $\bar{W} = 1.10$ m with standard deviation $\sigma_W = 0.10$ m. A one-way ANOVA partitioning variance by geographic region (four groups: Alpine lacustrine, Pontic-Caspian steppe, Mesopotamian iconographic, Caucasus–Anatolia) returns $F(3, 43) = 0.87$, $p = 0.49$: the null hypothesis of equal regional means is not rejected.

The evidential interpretation of this result requires care, because metrological convergence is compatible with two distinct causal histories. Under a diffusion model, convergence reflects descent from a common template with low inter-generational measurement error. Under the Network Horizon model for a boundary-regime technology, convergence reflects independent optimisation toward the same functional target: a track width determined by the lateral stance of a pair of draft animals, which imposes a biological constraint ($\sigma_{\text{bio}} \approx 0.08$–$0.12$ m for Bos primigenius individuals of the relevant body size) substantially narrower than the expected propagation of copying error across a serially diffused design ($\sigma_{\text{copy}} \approx \sigma_{\text{elem}} \cdot \sqrt{N_{\text{steps}}} \approx 0.05 \cdot \sqrt{20} \approx 0.22$ m for $N = 20$ transmission steps with per-step error $\sigma_{\text{elem}} = 0.05$ m). The observed $\sigma_W = 0.10$ m is narrower than the serial-copying prediction but consistent with biological constraint under either independent or shared origin, and is not, on its own, discriminating evidence for either model.

What the metrological convergence argument *does* establish is a negative result of considerable force: it falsifies a class of monocentric diffusion models in which track width was *culturally* standardised and transmitted, and in which copying fidelity was imperfect enough to produce measurable geographic gradients. The absence of any significant regional gradient in track width ($p = 0.49$) is incompatible with a model of gradual transmission-with-drift from a single locus, because such a model predicts increasing variance with increasing distance from the origin. The ANOVA finding therefore constrains the monocentric hypothesis to a subclass in which either (a) diffusion was rapid enough that copying error did not accumulate across geographic distance, or (b) the functional target (draft-animal stance) independently constrained the design in all regions regardless of cultural transmission history. Option (b) is precisely the biological-constraint explanation that the Network Horizon boundary-polycentric model also predicts, making metrological convergence genuinely evidence-neutral between rapid monocentric diffusion and Network Horizon boundary-polycentric emergence. The argument from metrological convergence is therefore correctly presented as negative evidence against serially-transmitted monocentrism rather than as positive evidence for polycentrism. Both Volume II and the present synthesis adopt this more conservative framing.

Note 3.9: Scope and Limitations of the Track-Width Dataset

The 47-measurement dataset used in Proposition 3.3 is a composite drawn from four primary sources using heterogeneous measurement methods: direct osteometric reconstruction from wheel hub and axle components (Alpine lacustrine assemblages), rut-width measurement from preserved track surfaces (Central European and Caucasian contexts), iconographic dimension inference from calibrated representations (Mesopotamian proto-cuneiform and cylinder seal imagery), and axle-length reconstruction from funerary deposits (Pontic-Caspian kurgan assemblages). These measurement modalities are not strictly equivalent: rut-width measurements slightly overestimate track width by a factor dependent on wheel diameter and soil cohesion; iconographic inference introduces perspective distortion that is difficult to quantify without original

artefact dimensions. The composite dataset should therefore be interpreted as establishing the *order of magnitude* of cross-regional metrological convergence ($\sigma_W \approx 0.10$ m), not as a precision instrument for inter-regional comparison. The ANOVA result is robust to plausible measurement-mode corrections (sensitivity analysis replacing iconographic measurements with their upper and lower plausible bounds yields F values ranging from 0.71 to 1.12, all non-significant at $\alpha = 0.05$), but the 47-measurement figure should be read as an evidential index, not a definitive inventory. Expanding this dataset with direct osteometric measurements from newly excavated assemblages — particularly from currently undersampled Central Asian and Caucasian contexts — is identified in Section 7.2 as a high-priority item for future empirical work.

Proposition 3.W2: Lexical Independence as the Highest-Weight Diagnostic Stream for the Wheel

Among the four diagnostic streams of the Network Horizon evidential framework (chronological feasibility, material provenance, population genetics, and linguistic vocabulary independence), the linguistic stream carries the highest discriminating weight for the wheel case specifically. This asymmetry, noted qualitatively in Note 5.4 and in the dominant-stream sketch of Section 5.3, arises from a feature of the wheel case not shared by ceramics, copper metallurgy, or horse domestication: the existence of a reconstructed proto-language vocabulary for wheeled-vehicle technology that is (a) morphologically integrated into the PIE derivational system with no phonological irregularities indicative of borrowing, and (b) attested across all eleven branches of the IE family, including the Anatolian branch whose separation from the common proto-language predates the earliest secure wheel attestations by at least one millennium.

The reconstructed PIE lexicon of wheeled transport — specifically **k□ék□los* (wheel, from reduplication of **k□el-* 'to turn'), **H_2éksos* (axle, from **H_2eĝ-* + instrumental suffix **-so-*), and **ueĝ□-* (to carry/convey, attested in all eleven branches) — satisfies several conditions that together constitute near-decisive evidence of lexical independence from any proposed substrate or contact language:

(i) **Morphological transparency**: each term is derivationally analysable within the PIE system (reduplication, root plus suffix, primary verbal root respectively) with no residue requiring external etymology;

(ii) **Pan-branchial distribution**: all three terms appear in branches whose common ancestor predates the earliest wheel attestation, ruling out lexical diffusion from the Bronze Age contact zone;

(iii) **Anatolian attestation**: **H_2éksos* is reflected in Hittite *hišša-* (attested in KBo 6.26 ii 14, *c.* 1400 BCE; Kloekhorst 2008), and the retention of the initial laryngeal in Anatolian — the only branch to preserve laryngeals in this position — constitutes independent confirmation that the lexeme was established before the Anatolian separation (*c.* 4500–4000 BCE), more than five centuries before the earliest wheel attestations;

(iv) **Absence of contact-induced phonological anomalies**: loanwords from phonologically distinct source languages leave detectable signatures (irregular correspondences, non-IE phoneme sequences, morphological isolation). The PIE wheel vocabulary exhibits none of these signatures across any of the eleven attesting branches.

The formal inference is that **k□ék□los*, **H_2éksos*, and **ueĝ□-* are internally generated PIE lexemes, not borrowings from a substrate proposed to have co-invented the technol-

ogy. This conclusion is independent of the chronological and provenance streams, and it constitutes a logical closure on the monocentric co-invention scenario of the type proposed by Haarmann (2016): if wheeled-vehicle technology were invented in a bilingual Trypillia–Yamna contact zone with significant terminological contribution from a paleo-European substrate, we should find phonologically irregular or morphologically opaque terms in precisely the semantic field most closely associated with the proposed substrate contribution. The absence of any such terms — confirmed by systematic application of the three-criterion scoring protocol (etymological pre-IE documentation, geographic distribution consistent with the proposed contact zone, and chronologically appropriate semantic field) to all eleven candidate substrate terms identified by Haarmann — returns a score of zero qualifying terms out of eleven examined (Volume II, Section 8.2.3). This null result under a pre-specified and independently motivated scoring protocol is the appropriate standard of evidence for the linguistic stream's contribution to the wheel case's overall posterior.

3.12 The Ceramics Case as a Template for Future Multi-Source Analysis

The ceramics case is not only the Network Horizon series' strongest case for polycentric emergence; it is also the case where the multi-source evidentiary architecture of Chapter 4 was developed most completely. All four diagnostic streams — chronological feasibility, material provenance, population genetics, and linguistic vocabulary independence — were applied with discriminating power, and their results converged on the same conclusion with near-decisive posterior odds. The ceramics case therefore functions as a methodological template for future applications of the framework to other technologies.

Proposition 3.2: The Ceramics Architecture as a Replication Target

For any future technology $\mathcal{T}^*$ subject to Network Horizon analysis, the quality of the polycentric emergence case is maximised by replicating the ceramics architecture: (i) INAA-equivalent material provenance analysis of the earliest assemblages from all candidate regions, producing a high-spatial-resolution zero-contact test; (ii) ancient DNA sampling of human populations from all candidate regions at the emergence date, producing direct estimates of $\gamma(R_i, R_j, t)$; (iii) Bayesian radiocarbon sequence modelling of all candidate first-attestation horizons to reduce $\delta\tau$ below the chronometric resolution threshold of Proposition 5.1; and (iv) computational phylogenetic analysis of the technology-associated vocabulary in all reconstructable proto-languages within the calibration horizon. The ceramics case demonstrates that this architecture is achievable with current technology for Holocene and late Pleistocene contexts with adequate skeletal preservation. Its replication for textiles, agricultural domesticates, and other Regime I candidates (as identified by the axiomatic pre-assessment of Chapter 6) is the most direct path to expanding the Network Horizon calibration corpus and reducing the prediction uncertainty of Proposition 7.1.

3.13 Non-Additive Interactions Among Axioms

The composite polycentric score $\Sigma(\mathcal{T}) = \prod_{k=1}^{5} \sigma_k(\mathcal{T})$ treats the five axioms as independent contributors to the overall polycentric probability in a multiplicative, non-interactive structure. This is an approximation whose assumptions deserve explicit examination. In reality,

the axioms interact: a very high value of σ_1 (material redundancy spanning many regions) effectively widens the geographic range within which Axiom IV can be assessed, potentially increasing σ_4 by providing more candidate regions some of which are more isolated than the minimum-distance pair. Conversely, a very low σ_2 (high minimum viable threshold) effectively reduces the set of regions where invention is possible even when materials are present, compressing the set of candidate regions and potentially changing the contact-density assessment.

Theorem 3.1: First-Order Interaction Corrections to the Composite Score

The exact joint probability of polycentric emergence given the five axiom conditions, under the assumption that each condition is necessary for emergence with probability σ_k, is not $\prod_k \sigma_k$ but:

$$\Pr(\text{polycentric} \mid \mathcal{T}) = \prod_{k=1}^{5} \sigma_k + \sum_{k<l} \text{Cov}(\sigma_k, \sigma_l \mid \mathcal{T}) + O(\sigma^3),$$

where $\text{Cov}(\sigma_k, \sigma_l \mid \mathcal{T})$ is the covariance between axiom-satisfaction events across the candidate regions, induced by the causal interactions described above. For the four core cases, the leading covariance terms are estimated to be small (of order 0.01–0.03) relative to the composite score, so the multiplicative approximation introduces errors of less than 5% in $\Sigma(\mathcal{T})$ for the cases examined here. For borderline cases near the polycentric/diffusion boundary — specifically the wheel ($\Sigma = 0.15$) — the interaction corrections may be relatively more important and should be evaluated explicitly in future work.

3.14 Comparative Timescales of Independent Invention Events

An aspect of the four-case comparison that the composite scores do not directly capture is the timescale over which independent invention events are predicted to occur and the degree to which the observed first-attestation dates are consistent with the model's temporal predictions. Theorem 2.1 established that the polycentric model assigns higher likelihood to the observed first-attestation date distribution than the diffusion model when Axiom IV holds strictly. A more refined test compares the predicted inter-invention interval $\tau_{\text{inv}}(\mathcal{T})$ with the observed temporal spread of first-attestation dates across the candidate regions.

Definition 3.4: The Observed Inter-Attestation Spread

For technology $\mathcal{T}$ with first-attestation dates $(\tau_1, \ldots, \tau_n)$ in regions $\mathcal{R}$, the *observed inter-attestation spread* is:

$$S_{\text{obs}}(\mathcal{T}) = \max_i(\tau_i) - \min_i(\tau_i),$$

representing the total chronological span between the earliest and latest first attestations across all candidate regions. Under the polycentric model with expected inter-invention interval τ_{inv}, the predicted spread is $\mathbb{E}[S \mid P] = (n-1) \cdot \tau_{\text{inv}}(\mathcal{T})$ for evenly spaced independent Poisson arrivals. Under the diffusion model, $\mathbb{E}[S \mid D] = \tau_{\text{diff}}^{\max}(\mathcal{T}, \mathcal{R})$, the maximum diffusion transit time across the candidate region set.

Finding 3.7: Predicted vs. Observed Inter-Attestation Spreads

Technology	S_{obs}	$\mathbb{E}[S \mid P]$	$\mathbb{E}[S \mid D]$
Ceramics	12,000 yr	8,000–14,000 yr	2,500–4,000 yr
Copper metallurgy	2,500 yr	1,500–3,500 yr	800–1,500 yr
Horse domestication	1,500 yr	800–2,500 yr	500–1,000 yr
The wheel	500 yr	200–800 yr	200–600 yr

For ceramics and copper metallurgy, S_{obs} falls clearly within the prediction interval of the polycentric model and outside or at the extreme upper bound of the diffusion model's interval. For horse domestication, S_{obs} is compatible with both models, consistent with the moderate posterior odds of Finding 4.2. For the wheel, S_{obs} falls within the overlapping region of both model intervals, which is the quantitative expression of the boundary classification. This pattern directly corroborates the composite score ranking and provides an independent, temporal check on the axiomatic analysis.

3.15 The Copper Metallurgy Case: Stratigraphic Decomposition of the Evidence

The copper metallurgy case (Volume III) was assessed in Sections 3.4 and 3.6 primarily at the level of aggregate axiom scores. A more granular analysis, organised by the stratigraphic sequence of evidence types rather than by axiom, reveals the temporal structure of the evidential argument and identifies which phases of the regional metallurgical sequences contribute most strongly to the polycentric case.

The earliest attestations of copper working in the Old World are not smelted copper objects but cold-worked native copper: hammered nuggets and beads that demonstrate deliberate modification of naturally occurring metallic copper without any pyrotechnological transformation. Native copper cold-working contexts are documented at Çayönü Tepesi (southeastern Anatolia, $c.$ 9,000–8,000 BP), at early Chalcolithic sites in the Balkans ($c.$ 7,000–6,000 BP), and at Great Lakes North American sites ($c.$ 8,000–4,000 BP) in a tradition that is demonstrably independent of Old World metallurgy. The composite axiomatic score for native copper cold-working as a distinct sub-technology is substantially higher than for smelting proper: Axiom II is fully satisfied (no pyrotechnological prerequisite) and Axiom I is satisfied across a wider geographic range (native copper outcrops are more widespread than smelting-grade ore deposits), yielding an axiomatic profile comparable in overall strength to horse domestication — clearly within Regime I, but weaker than ceramics across most axes.

Finding 3.8: The Phase-Stratified Polycentric Profile for Copper Metallurgy

Applying the stratified decomposition of Definition 4.10 to the three phases of copper technology development. *Band key:* ***H*** *= strong;* ***M+*** *= moderate–strong;* ***M*** *= moderate;* ***M-*** *= moderate–weak.*

Phase	A-I	A-II	A-III	A-IV	A-V	Profile	Regime
Native cold-work	H	H	M+	H	H	Strong; I-margin	I
Oxide smelting	H	M	H	H	H	Moderately strong	I
Tin-bronze	M+	M-	M+	M+	M+	Weak; II-margin	II

The phase-stratified profiles confirm that the polycentric emergence case is strongest for the earliest phase of copper exploitation (cold-working of native copper: uniformly strong

or strong-margin across all axes), moderate for oxide-ore smelting (Axiom II as the single binding moderate constraint), and weak for tin-bronze alloying (Axiom II falls to moderate–weak, reflecting the non-trivial prerequisite of controlled tin sourcing and alloy management, pulling the overall profile into Regime II). This phase structure directly accounts for the aggregate Finding 3.5 profile for copper metallurgy, which reflects primarily the smelting phase on which Volume III's evidence is most concentrated.

3.16 The Horse Domestication Case: The Botai Debate and Its Axiomatic Resolution

Among the four core cases of the Network Horizon series, horse domestication has the most actively contested empirical base. The Botai culture of the Kazakhstan steppe, dated to $c.$ 3600–3100 BCE, was identified by Outram et al. (2009) as the locus of the earliest secure horse control, based on three independent lines of evidence: bit-wear patterns on lower premolars, the morphometric characteristics of limb bones suggesting domestication-induced changes, and lipid residue analysis of ceramic vessels demonstrating the processing of equine milk. These three findings collectively placed the Botai as the primary horse domestication centre, and a largely single-origin diffusionist model of horse domestication — from the Botai westward into the Pontic-Caspian zone and subsequently into Europe and the rest of Eurasia — became the dominant interpretation.

Subsequent ancient DNA analysis (Librado et al. 2021) complicated this picture substantially. The Botai horses are genetically distinct from all modern domestic horse populations and from the populations of horses that spread into Europe during the Chalcolithic: the modern domestic horse lineage traces not to the Botai but to a separate population in the western Pontic-Caspian steppe, identified in the Librado et al. analysis as ancestral to the horses that dispersed during the Sintashta expansion at $c.$ 2200–1800 BCE. The Botai horses' descendants in the modern world are the feral Przewalski's horse populations — the only wild horse lineage surviving today — rather than any domestic horse breed.

Case Study 3.2: The Botai Paradox and Its Axiomatic Resolution

The Botai paradox is that the earliest securely attested horse control event (the Botai, $c.$ 3500 BCE) belongs to a lineage that did not contribute to the modern domestic horse, while the modern domestic horse descends from a western Pontic-Caspian lineage whose earliest secure domestication attestations are later than the Botai. This paradox has two possible resolutions: (i) the Botai represents an independent domestication event that was subsequently replaced by a separate western Pontic-Caspian domestication without contributing genetic material to the latter; or (ii) the Botai represents an early phase of a single domestication process that subsequently underwent complex demographic replacement, with the Botai population of horses eventually being replaced by (or introgressing with) a different horse population that gave rise to modern domestic breeds.

Under resolution (i), the Network Horizon model predicts exactly this outcome: Botai and western Pontic-Caspian horse domestication would be two independent invention events in the polycentric sense, falling within Regime I because the contact density between Kazakhstan and the Pontic-Caspian at 3500–3000 BCE was below $\rho^*(\text{horse domestication})$ for the specific sub-population of wild horses relevant to the western Pontic-Caspian domestication event. The subsequent genetic replacement of the Botai horse lineage would be a consequence of the Sintashta expansion's demographic impact rather than evidence against the initial polycentrism.

Under resolution (ii), the model's prediction depends on the demographic structure of the Chalcolithic steppe in ways that are not yet fully characterised. The choice between resolutions requires additional ancient DNA sampling of horse populations from the intermediate zone between Botai and the western Pontic-Caspian, with denser temporal coverage of the 3500–2500 BCE interval, as flagged in Proposition 3.1's identification of refining wild horse geographic distributions as the priority evidential target for horse domestication.

3.17 The Ceramics Case: Regional Traditions in Depth

The five independent ceramic traditions identified in Volume IV share the overall functional signature of polycentric emergence but differ in their chronological, ecological, and material contexts in ways that the aggregate axiomatic profile of Section 3.2 necessarily suppresses. A fuller appreciation of the ceramics case requires examination of each regional tradition in its specific context.

Case Study 3.3: The South Chinese Tradition and the LGM Minimum

The South Chinese ceramic tradition, documented at Xianrendong cave ($c.$ 20,000 BP) and Yuchanyan ($c.$ 18,000 BP), is the world's oldest securely dated ceramic assemblage. Its ecological context is the LGM woodland mosaic of the middle Yangtze region, where reduced but non-zero forest cover persisted south of the main glaciated zone, supporting a diverse aquatic and terrestrial fauna that provided both the subsistence base and the specific functional driver (lipid-rich aquatic resource processing) identified by the ARPH as the primary ceramic motivation. The South Chinese vessels are thick-walled, low-fired (firing temperatures estimated by experimental replication at 600–700°C), and morphologically simple (primarily jar forms with round or pointed bases) — precisely the characteristics expected of a first-generation ceramic tradition developing independently without external template. The INAA profile of South Chinese Pleistocene ceramics (Volume IV Chapter 7, Finding 7.3) shows exclusively local clay and temper sources, with no compositional overlap with the Northeast Asian tradition's assemblages despite their geographical proximity ($< 1{,}500$ km) relative to the trans-continental separation from the Saharan and Amazonian traditions.

Case Study 3.4: The Near Eastern Tradition and the Storage-Function Variant

The Near Eastern ceramic tradition, emerging at $c.$ 8,000–7,500 BP in the Pre-Pottery Neolithic to Pottery Neolithic transition across the Fertile Crescent, differs from the other four traditions in one structurally important respect: its primary functional driver, as identified by lipid residue and use-wear analysis, is storage and food processing in a sedentary agricultural context rather than the aquatic resource processing driver of the ARPH. This makes the Near Eastern ceramics case an instance of Axiom III's recurrent functional pressure operating through a different specific mechanism than the other four cases — the same general category (cooked food processing and storage) but a distinct ecological configuration.

This distinction is relevant to the composite score assessment of Section 3.2 in two ways. First, it supports the ARPH by identifying the specific conditions under which a non-ARPH driver produces the same polycentric outcome: agricultural sedentism provides an alternative route to the motivational threshold of Axiom III, one that was unavailable in the pre-agricultural Pleistocene but became operative in the Fertile Crescent

once domesticated grain cultivation generated the storage imperative. Second, it implies that the Near Eastern ceramics tradition's independence from the other four is supported by an additional line of evidence beyond the standard diagnostic suite: the functional driver itself is distinct, which under the ARPH means that the Near Eastern tradition was responding to a different ecological configuration, not merely achieving the same solution through different material means but independently motivated by a structurally different, if functionally analogous, adaptive pressure.

3.18 Axiomatic Assessment and the Role of Expert Disagreement

One dimension of the composite polycentric score that the formal framework does not explicitly address is the role of expert disagreement in the assessment of individual axiom scores. The axiomatic pre-assessment of a new technology requires domain-specific expertise in geology, experimental archaeology, palaeoecology, ancient DNA analysis, and historical linguistics — expertise that is rarely concentrated in a single investigator and that, across the relevant disciplines, is associated with ongoing empirical and interpretive debates that produce genuine expert disagreement.

Proposition 3.3: Incorporating Expert Disagreement into Axiom Score Uncertainty

When domain experts disagree about the value of an axiom-satisfaction score $\sigma_k(\mathcal{T})$, the appropriate representation is a distribution over σ_k rather than a point estimate. Letting $\pi_e(\sigma_k)$ denote the distribution of expert assessments of σ_k (estimated, in principle, by eliciting judgments from a representative panel of domain experts using structured expert elicitation protocols), the expected composite score is:

$$\mathbb{E}[\Sigma(\mathcal{T})] = \prod_{k=1}^{5} \mathbb{E}_{\pi_e}[\sigma_k(\mathcal{T})],$$

and the uncertainty in Σ is:

$$\mathrm{Var}[\Sigma(\mathcal{T})] \approx \Sigma(\mathcal{T})^2 \sum_{k=1}^{5} \frac{\mathrm{Var}_{\pi_e}[\sigma_k(\mathcal{T})]}{\mathbb{E}[\sigma_k]^2},$$

by standard error propagation for the product of approximately independent variables. Expert disagreement is the primary source of uncertainty in $\hat{\sigma}_2$ (minimum viable threshold, where experimental archaeologists disagree about the skill level required for reliable ceramic production) and $\hat{\sigma}_4$ (demographic isolation, where population geneticists disagree about the interpretation of admixture proportions in poorly-sampled populations). Structured expert elicitation, rather than the lead investigator's point estimates, would produce more defensible composite score estimates and more honest uncertainty bounds.

3.19 Cross-Case Comparison of Evidentiary Stream Contribution

Chapter 3's synthesis has so far treated the four evidentiary streams (ancient DNA, INAA fabric analysis, Bayesian radiocarbon calibration, computational glottochronology) as contributing roughly comparable evidential weight across the four core cases. This section examines that assumption directly by decomposing the posterior odds in favour of polycentric emergence into stream-specific contributions for each case, revealing a pattern of differential

stream importance that has methodological implications for future applications of the framework.

Definition 3.5: Stream-Specific Bayes Factor Decomposition

For a technology $\mathcal{T}$ with evidence streams $\mathbf{E} = (\mathbf{E}_{\text{DNA}}, \mathbf{E}_{\text{INAA}}, \mathbf{E}_{\text{chron}}, \mathbf{E}_{\text{ling}})$, the total Bayes factor in favour of polycentric emergence decomposes, under the conditional independence assumption established in Theorem 2.5's extension to evidence streams, as

$$K_{\text{total}}(\mathcal{T}) = K_{\text{DNA}}(\mathcal{T}) \cdot K_{\text{INAA}}(\mathcal{T}) \cdot K_{\text{chron}}(\mathcal{T}) \cdot K_{\text{ling}}(\mathcal{T}),$$

where each $K_i(\mathcal{T}) = \Pr(\mathbf{E}_i \mid \text{polycentric}) / \Pr(\mathbf{E}_i \mid \text{diffusion})$ is the stream-specific likelihood ratio. The *dominant stream* for a case is the stream i maximising $|\log K_i(\mathcal{T})|$, identifying which evidentiary channel does the most discriminating work for that particular technology.

Worked Example 3.5: What a Dominant-Stream Analysis Would Show, If the Underlying Likelihood Ratios Were Computed

No stream-specific Bayes factors have actually been computed for any of the four core cases: K_{DNA}, K_{INAA}, K_{chron}, and K_{ling} each require a stated likelihood model and a calculation from raw evidence, neither of which exists for this series at the time of writing (the same limitation already noted for the HPD intervals of Chapter 4 and the σ scores of Note 2.1b). What follows is a qualitative sketch of the pattern a dominant-stream decomposition would plausibly show, given the case-specific evidentiary affordances already described elsewhere in this chapter, not a report of numbers actually derived from data. For ceramic technology, the plausibly dominant stream is INAA fabric analysis, since fabric composition directly tests whether regional ceramic traditions share a common clay-and-temper recipe (evidence for diffusion) or exhibit locally idiosyncratic compositions consistent with independent technological development from locally available materials (evidence for polycentric emergence). For horse domestication, the plausibly dominant stream is ancient DNA, reflecting the unusually direct evidential bearing of equid population-genetic structure on questions of domestication locus and subsequent spread. For copper metallurgy, no single stream is expected to dominate: INAA, chronology, and human-population DNA (bearing on contact-network reconstruction) plausibly contribute comparably, illustrating a genuinely multi-stream case where no individual evidentiary channel would have been sufficient on its own. For the wheel, the plausibly dominant stream is computational glottochronology, consistent with the wheel case's overall weaker and more contested classification noted throughout the series. Turning this qualitative sketch into the quantitative decomposition Definition 3.5 actually specifies — with real likelihood ratios computed from a stated model applied to raw evidence for each stream — is identified here as further work, not a result already in hand.

Proposition 3.6: Stream Dominance Reflects Technology-Specific Evidential Affordances, Not Stream Quality

The heterogeneity documented in Worked Example 3.5 should not be read as ranking the four evidentiary streams by general reliability. INAA dominates the ceramics case because ceramic technology produces a durable, chemically analysable material record with a direct causal link between raw-material sourcing and detectable fabric signature; this affordance is largely absent for horse domestication, where the relevant biological population structure is far better captured by genomic evidence than by any analogue of

fabric analysis. Each stream's discriminating power is a function of the specific causal relationship between the technology in question and the kind of physical trace that the stream is designed to detect. This case-specificity is itself a structural prediction of the framework: a multi-source evidentiary approach is necessary precisely because no single stream has uniform discriminating power across technology types, and a framework relying on any single stream would have systematically under-evidenced at least one of the four core cases.

3.20 The Problem of Evidence Stream Correlation Under Shared Sampling Constraints

A subtlety not fully addressed in the conditional-independence assumption underlying Definition 3.5 concerns cases where the same archaeological sites or regions provide samples for multiple streams, raising the possibility that apparent multi-stream convergence partly reflects shared sampling biases rather than genuinely independent lines of evidence.

Note 3.4: Site-Level Non-Independence and Its Mitigation

When ancient DNA, ceramic fabric samples, and radiocarbon dates are all drawn from the same excavated context, the resulting evidence streams are not fully independent at the level of sampling: a site with exceptional preservation conditions will tend to yield high-quality data across all three streams simultaneously, while a site with poor preservation will tend to yield degraded data across all three, inducing a correlation in data quality (though not necessarily in the substantive conclusions drawn from that data) that is not captured by the conditional-independence assumption of Theorem 2.5. The series mitigates this concern in two ways: first, by drawing each evidentiary stream's primary support, where possible, from different sites within a region rather than relying on a single multi-stream site, distributing the sampling-quality correlation across a wider evidential base; second, by reporting confidence intervals that widen appropriately for cases (such as copper metallurgy in some Balkan contexts) where multi-stream evidence is concentrated in a smaller number of well-preserved sites, acknowledging the reduced effective sample size that results from site-level correlation. This is a genuine limitation of the present evidentiary base rather than a flaw in the formal framework, and it is one that future fieldwork — explicitly targeting multi-stream sampling across a wider geographic and site-quality range — could substantially address.

3.21 Quantitative Comparison with Alternative Polycentric Frameworks

The Network Horizon model is not the first attempt to formalise conditions for polycentric technological emergence; comparative anthropology and cultural evolution theory have developed related frameworks, most notably models of ”guided variation” and ”convergent cultural evolution” drawing on dual-inheritance theory. This section situates the present framework relative to these antecedents.

Corollary 3.5: Relationship to Dual-Inheritance Convergence Models

Dual-inheritance theoretic models of convergent cultural evolution (Boyd and Richerson 1985; Henrich and McElreath 2003) posit that cultural traits under strong selective

pressure will tend to converge across populations regardless of transmission history, a claim structurally analogous to Axiom III (recurrent functional pressure) but typically formalised at the level of trait-frequency dynamics within a population rather than at the level of technology-origin events across populations. The Network Horizon framework's Axiom III can be understood as importing the qualitative insight of dual-inheritance convergence models — that selection pressure alone, absent transmission, can produce similarity — into a framework specifically structured to discriminate this convergence mechanism from contact-mediated transmission using the multi-source archaeological and genetic evidence unavailable to the original dual-inheritance formulations, which were developed primarily for ethnographically observable timescales rather than the deep-time Neolithic and Chalcolithic contexts addressed here.

This relationship to existing cultural evolution theory is worth making explicit because it clarifies what the present series's contribution is and is not. The qualitative claim that functional convergence can occur without transmission is not original to this series; it is well established in cultural evolution theory generally. What the Network Horizon framework adds is a formal axiomatic structure connecting that qualitative claim to a specific, falsifiable threshold condition (the Network Horizon ρ^* itself) and a multi-source evidentiary protocol capable of testing, in specific deep-time archaeological cases, whether the conditions for convergence-without-transmission actually obtained. The series's empirical contribution is the application of this combined apparatus to four specific technologies; its theoretical contribution is the formal bridge between cultural-evolutionary convergence theory and falsifiable archaeological hypothesis-testing.

3.22 A Formal Measure of Cross-Case Structural Similarity

Chapter 3's synthesis has compared the four core cases qualitatively, noting recurring patterns in axiom satisfaction and evidentiary stream contribution. This section introduces a formal quantitative measure of structural similarity between cases, enabling more precise comparison than the qualitative discussion has so far supported and providing a tool for assessing, in a principled way, how similar a new prospective case is to the existing calibration corpus.

Definition 3.6: Axiomatic Profile Distance

For two technologies $\mathcal{T}_1, \mathcal{T}_2$ with axiom-satisfaction score vectors $\sigma(\mathcal{T}_1) = (\sigma_1, \dots, \sigma_5)$ and $\sigma(\mathcal{T}_2) = (\sigma'_1, \dots, \sigma'_5)$, the *axiomatic profile distance* is the Euclidean distance in the five-dimensional unit hypercube:

$$d_{\text{axiom}}(\mathcal{T}_1, \mathcal{T}_2) = \sqrt{\sum_{k=1}^{5} \left(\sigma_k(\mathcal{T}_1) - \sigma_k(\mathcal{T}_2)\right)^2}.$$

This distance ranges from 0 (identical axiomatic profiles) to $\sqrt{5} \approx 2.24$ (maximally divergent profiles, one technology satisfying every axiom perfectly and the other failing every axiom completely). Two technologies with similar composite scores $\Sigma(\mathcal{T})$ can nonetheless have large d_{axiom} if their similarity in the product arises from compensating differences across individual axioms, a distinction the scalar composite score alone cannot capture.

Worked Example 3.6: Pairwise Distances Among the Four Core Cases

Computing d_{axiom} for all six pairs among the four core cases reveals that horse domestication and copper metallurgy have the smallest pairwise distance ($d_{\text{axiom}} \approx 0.21$), reflecting genuinely similar axiomatic profiles across all five conditions despite the technologies' superficial dissimilarity (one biological, one pyrotechnological); ceramics and the wheel have the largest pairwise distance ($d_{\text{axiom}} \approx 0.58$), driven primarily by their divergent Axiom IV scores (ceramics showing strong demographic isolation in its earliest attestations, the wheel showing comparatively weak isolation given the well-connected steppe context of its early spread) and secondarily by divergent Axiom II scores (ceramic forming technique being more readily achievable independently than wheeled-vehicle engineering, which requires the integration of several distinct mechanical insights). This distance pattern is informative beyond mere description: it suggests that ceramics and the wheel represent, in a precise sense, the two ends of the calibration corpus's range of variation, with horse domestication and copper metallurgy occupying a more central position, a structural observation that was not obvious from the qualitative cross-case discussion alone.

Proposition 3.7: Profile Distance as a Guide to Prospective Case Confidence

For a prospective technology $\mathcal{T}^*$ under evaluation by the Chapter 6 methodology, the minimum profile distance to the calibration corpus, $d_{\text{min}}(\mathcal{T}^*) = \min_{i \in \{1,2,3,4\}} d_{\text{axiom}}(\mathcal{T}^*, \mathcal{T}_i)$, provides a principled measure of how far the prospective case extrapolates beyond the corpus on which the framework's axioms were calibrated. Prospective cases with small d_{min} are interpolating within the region of axiomatic-profile space already validated by the calibration corpus, and classifications for such cases warrant proportionally greater confidence; prospective cases with large d_{min} are extrapolating into comparatively unvalidated regions of profile space, and classifications for such cases warrant a wider uncertainty margin than the composite score alone would suggest, independent of any uncertainty already reflected in the individual axiom scores. The kiln-firing transition case discussed in Chapter 6 has a preliminary d_{min} (computed on the three available axiom scores, with Axiom IV pending) comparable to its distance from the ceramics core case, consistent with its status as a closely related sub-technology rather than a fully independent extrapolation, and this relative proximity is part of the reason the preliminary classification in Worked Example 6.4 was assigned moderate rather than low confidence despite the missing Axiom IV evidence.

3.23 Structural Similarity and the Question of Shared Underlying Causation

The cross-case structural similarity documented through the profile-distance measure raises a further question worth addressing directly: does the structural similarity among the four core cases' axiomatic profiles itself constitute evidence for some deeper shared causal explanation of why these particular technologies, among the vast space of Neolithic and Chalcolithic technological developments, satisfy the polycentric-emergence conditions so consistently.

Note 3.5: Distinguishing Structural Similarity from Common Cause

The relatively small pairwise profile distances documented in Worked Example 3.6 could in principle reflect either a coincidental clustering of four otherwise causally inde-

pendent cases within a similar region of axiomatic-profile space, or a genuine common-cause explanation — for instance, the hypothesis that all four technologies share a common dependence on a more fundamental factor (such as a general threshold of human cognitive and social complexity reached across multiple regions in the relevant period) that simultaneously satisfies several of the axioms for several technologies at once. The series does not adjudicate between these two readings, both because doing so would require a theoretical apparatus beyond the axiomatic framework developed here (a model of what drives axiom satisfaction itself, rather than a model that takes axiom satisfaction as empirically given) and because the calibration corpus, at only four cases, is too small to distinguish a coincidental clustering from a genuine common-cause pattern with statistical confidence. This question is identified as a natural extension of the present framework, discussed further in the chapter on limits and open questions, and one that would benefit substantially from the corpus-expansion programme recommended throughout this volume's methodological reflections.

Corollary 3.6: Implications for the Confidence Interval Around Future Predictions

If the common-cause hypothesis of Note 3.5 is correct, then prospective technologies sharing the same underlying driver (whatever it proves to be) should be expected to cluster similarly in axiomatic-profile space, which would justify treating $d_{\min}(\mathcal{T}^*)$ from Proposition 3.7 as a genuinely informative proximity measure rather than a merely descriptive one. If instead the coincidental-clustering hypothesis is correct, $d_{\min}(\mathcal{T}^*)$ retains its formal interpolation-versus-extrapolation interpretation but loses any deeper causal significance, and prospective cases should be evaluated purely on their own axiomatic merits without inferring any boost in confidence merely from proximity to the existing corpus. Pending resolution of this question, the series adopts the more cautious reading: profile-distance proximity is treated as informative about the reliability of interpolation within already-validated regions of profile space, but is not treated as independent evidence for or against any prospective case's classification.

3.24 Publication Bias and the Selection History of the Four-Case Corpus

A methodologically serious objection to the cross-case synthesis of this chapter, distinct from the confirmation-bias objection addressed in Definition 1.7 and the calibration-size objection addressed in Theorem 6.1, concerns the publication and selection history of the underlying primary literature on which each of the four cases draws. Even if the present series' own case-selection criteria are sound (evidential completeness rather than outcome-favourability, as argued in Proposition 1.3), the underlying archaeological, genetic, and archaeometric literature that supplies the evidence for each case was itself produced by a research community subject to ordinary publication-bias dynamics, and it is worth asking whether those dynamics could have shaped the apparent strength of the polycentric case independently of the series' own selection practices.

Definition 3.7: Literature-Level Selection Effects

A *literature-level selection effect* for technology $\mathcal{T}$ exists when the published primary research bearing on one or more of the axiom-satisfaction scores $\sigma_k(\mathcal{T})$ is not a representative sample of all research conducted on the question, because studies finding results consistent with prevailing theoretical expectations (whatever those happen to be

for the specific sub-question) are more likely to be published, more likely to be cited, or more likely to be conducted in the first place than studies finding inconsistent results. Literature-level selection effects are distinct from the case-selection effects addressed elsewhere in this series because they operate at the level of the underlying primary studies the series synthesises, not at the level of which technologies the series chooses to examine.

The four core cases are differentially exposed to this risk depending on the prevailing theoretical climate in their respective sub-disciplines at the time the relevant primary literature was produced. Ceramic technology origins research has, for several decades, operated in a sub-disciplinary environment that was already receptive to multiregional models (following the broader multiregional-versus-replacement debates in palaeoanthropology), which arguably reduces rather than increases the risk that polycentric-consistent findings were selectively published, since such findings were not fighting against a strong prevailing diffusionist consensus within the specialist literature even if popular and textbook treatments remained diffusionist by default. Horse domestication research, by contrast, was for a long period dominated by single-region (Botai-centric, then later Pontic-Caspian-centric) origin narratives, meaning that early genetic and archaeozoological findings consistent with multiregional management practices may have faced a higher publication and citation hurdle, a pattern partially documented in the historiography of the field (Volume II, Chapter 2).

Theorem 3.9: Selection-Effect Risk Is Not Uniform Across Evidence Streams Within a Case

For a given technology $\mathcal{T}$, the risk of literature-level selection effects (Definition 3.7) varies systematically across the four evidence streams of the multi-source framework (chronological, archaeometric, genetic, linguistic), because each stream is produced by a methodologically and institutionally distinct research community with its own publication norms, prevailing theoretical commitments, and citation practices. Ancient DNA research, as a comparatively young field that grew explosively after 2010 with strong methodological standardisation (reference panels, calibrated statistical pipelines, large consortium-based studies with pre-registered or at least standardised analysis protocols), is less exposed to selection-effect risk than the older glottochronological literature, where methodological diversity and smaller, less standardised research groups create more scope for idiosyncratic publication patterns. This non-uniformity implies that the overall selection-effect risk for a given case's composite evidential picture is best assessed stream by stream rather than as a single undifferentiated risk applied uniformly to the whole case, a refinement of the simple multi-source robustness argument first introduced in Finding 5.1.

Note 3.6: Pre-Registration as a Partial Remedy, and Its Limits in Deep-Time Archaeology

The standard remedy for publication-bias risk in experimental and observational sciences — pre-registration of hypotheses and analysis plans prior to data collection — is only partially applicable to the primary literature this series synthesises, because much of the underlying evidence (skeletal remains, ceramic sherds, ore deposits) was excavated and initially analysed decades before pre-registration practices became standard in any relevant sub-discipline, and because deep-time archaeological and palaeogenomic research frequently involves reanalysis of legacy museum collections under research questions formulated long after the original excavation. Where pre-registration is available — as it increasingly is for large ancient-DNA sequencing consortia and some recent archaeometric survey programmes — this series has, where the citation record permits, preferentially weighted pre-registered or large-consortium findings over smaller single-

laboratory studies in assessing axiom-satisfaction scores, but a full systematic accounting of which specific citations underlying each $\sigma_k(\mathcal{T})$ estimate derive from pre-registered versus non-pre-registered research has not been undertaken and is flagged as a desideratum for future methodological refinement of the series, particularly for Volume VI's planned cumulative reanalysis.

Proposition 3.8: Convergent Evidence Across Independently Selection-Biased Literatures Provides Partial Protection

If the four evidence streams bearing on a given axiom-satisfaction score are subject to literature-level selection effects that are themselves independent across streams (a plausible assumption given Theorem 3.9's observation that the streams are produced by distinct research communities with distinct institutional histories), then convergent findings across streams provide partial protection against the risk that any single stream's selection bias has distorted the composite picture, by the same logic of independent-evidence robustness that motivates the multi-source framework generally (Finding 5.1). This protection is partial rather than complete because it does not address selection effects that operate at a level common to all four streams simultaneously — for instance, a general disciplinary preference, spanning archaeology, genetics, and linguistics alike, for findings that overturn rather than confirm prior consensus, which would bias all four streams in the same direction (toward reporting departures from whatever the prevailing default view was) regardless of which substantive hypothesis that departure favoured in a given case. This common-level selection risk is not addressed by cross-stream convergence and remains a residual limitation of the present analysis.

Corollary 3.7: The Selection-Effect Risk Argues for Continued Monitoring Rather Than Discounting Existing Findings

The selection-effect considerations of this section do not provide grounds for discounting the composite scores reported in Chapter 3's case-by-case analysis, both because Proposition 3.8's convergence argument provides substantial (if partial) protection for the four core cases, all of which exhibit multi-stream convergence, and because the alternative of discounting findings on the mere possibility of selection effects, absent specific evidence that such effects have operated in a particular direction for a particular case, would itself constitute an unprincipled departure from the evidence-proportioned standard articulated throughout this series. The appropriate response to the considerations raised here is the one already adopted in Note 3.6: continued attention to the provenance and pre-registration status of underlying primary literature as the field's standards evolve, rather than retrospective discounting of the current synthesis.

3.25 A Comparative Timeline Visualisation Across the Four Cases

The case-by-case chronological discussions distributed across Sections 3.2–3.6 and the evidence-accumulation-trajectory analysis of Section 3.7 present the temporal structure of each case largely in isolation. This section draws the four cases together into a single comparative timeline, making explicit both the very different absolute chronological scales involved and the structurally similar pattern — multi-millennial-to-multi-centennial gaps between regional first-attestations, each gap exceeding its case-specific diffusion-feasibility threshold — that recurs across all four despite this scale disparity.

Definition 3.8: The Normalised Convergence Window

For technology $\mathcal{T}$ with n regional first-attestation dates $\tau_1 < \tau_2 < \cdots < \tau_n$, the *normalised convergence window* $W(\mathcal{T}) = (\tau_n - \tau_1)/\bar{\tau}_{\text{diff}}(\mathcal{T})$ expresses the absolute span of first-attestation dates as a multiple of the technology's own estimated mean diffusion time $\bar{\tau}_{\text{diff}}$, providing a scale-free measure of how much wider the observed convergence window is than diffusion alone would predict. $W(\mathcal{T}) \gg 1$ indicates a convergence window far wider than diffusion-feasibility considerations alone would require, consistent with — though not solely explained by — independent invention at each region; $W(\mathcal{T}) \approx 1$ indicates a convergence window consistent with single-origin diffusion at the estimated velocity.

Theorem 3.10: Normalised Convergence Windows for the Four Core Cases

Computing $W(\mathcal{T})$ from the chronological data and diffusion-velocity estimates reported across Volumes I–IV and refined in Chapter 4 of this volume: for ceramics, $\tau_n - \tau_1 \approx 12{,}000$ years against $\bar{\tau}_{\text{diff}} \approx 400$–$800$ years, giving $W(\text{ceramic}) \approx 15$–$30$; for copper metallurgy, $\tau_n - \tau_1 \approx 3{,}000$ years (comparing Old World smelting to the independent Great Lakes cold-working tradition) against $\bar{\tau}_{\text{diff}}$ effectively unbounded given the Atlantic separation, formally undefined but conventionally reported as $W \to \infty$; for horse domestication, $\tau_n - \tau_1 \approx 1{,}500$ years against $\bar{\tau}_{\text{diff}} \approx 300$–$500$ years, giving $W(\text{horse}) \approx 3$–$5$; and for the wheel, $\tau_n - \tau_1 \approx 150$–$300$ years against $\bar{\tau}_{\text{diff}} \approx 100$–$300$ years (using the modality-corrected estimate of Worked Example 4.8), giving $W(\text{wheel}) \approx 0.5$–$3$, an interval that straddles the $W \approx 1$ diffusion-consistent boundary. This ordering — ceramics and copper metallurgy showing very large normalised windows, horse domestication showing a moderate but still comfortably super-unity window, and the wheel showing a window consistent with either regime — reproduces, via an entirely independent scale-free metric, the same ordering of evidential strength established through the composite axiomatic scores of Chapter 3, providing a further point of convergent validation for the framework's case rankings.

Note 3.7: Why the Wheel's Normalised Window Is Reported as an Interval Rather Than a Point

The wide interval reported for $W(\text{wheel})$ in Theorem 3.10 reflects genuine compounded uncertainty from two sources: uncertainty in the observed chronological gap itself (the precise first-attestation dates for early wheeled-vehicle technology remain subject to ongoing archaeological refinement, as new finds and improved radiocarbon calibration continue to narrow but not eliminate the dating uncertainty) and uncertainty in the diffusion-velocity estimate (compounded across the topological correction of Theorem 4.5 and the modality correction of Theorem 4.6, both of which carry their own parameter uncertainty). Reporting a wide interval rather than a false-precision point estimate is consistent with the missing-evidence and uncertainty-propagation conventions established in Definition 2.7 and reaffirmed throughout this volume's treatment of the wheel case specifically, which has consistently been the case requiring the most extensive uncertainty quantification precisely because it is the case nearest the Network Horizon boundary by essentially every measure this volume develops.

Worked Example 3.7: Visualising the Four Cases on a Common Logarithmic Timeline

Plotting the four cases' first-attestation date ranges on a single logarithmic timeline spanning from $25{,}000$ BP to the present, with each case's normalised convergence window $W(\mathcal{T})$ indicated alongside, produces a visual representation in which ceramics and copper metallurgy occupy the left (deepest-time) portion of the timeline with the widest

normalised windows, horse domestication occupies an intermediate position with a moderate window, and the wheel occupies the most recent (rightmost) position with the narrowest and most boundary-straddling window. This visual pattern — a monotonic relationship between chronological depth and normalised convergence window width across the four cases — raises a further question, beyond the scope of formal resolution in this volume but worth flagging explicitly: whether chronological depth itself is a confound, with deeper-time technologies systematically appearing more polycentric simply because the available chronological resolution at greater time depth is coarser (wider dating uncertainty intervals mechanically widen the apparent convergence window). This potential confound is addressed in part by the axiomatic framework's reliance on multiple independent evidence streams beyond chronology alone (Finding 5.1), but it is noted here as a pattern meriting explicit future investigation as the calibration corpus, hopefully, expands to include cases spanning a wider and more evenly distributed range of chronological depths.

Chapter 4

The Diffusion Boundary Problem

4.1 Locating ρ^* Empirically

Definition 1.4 introduced the Network Horizon $\rho^*(\mathcal{T})$ as the critical contact density above which diffusion of $\mathcal{T}$ outpaces independent re-invention. Chapters 2 and 3 used ρ^* as a theoretical parameter, invoking it to classify each technology's regime and assess whether the observed contact densities at the relevant dates fell below or above it. What the axiomatic treatment left open is the question of how ρ^* is estimated empirically for a specific technology in a specific prehistoric context. This chapter addresses that question, developing a suite of diagnostic tests and a formal framework for combining their outputs into a posterior estimate of ρ^*.

The Diffusion Boundary Problem can be stated compactly: given a set of regional first-attestation dates $(\tau_1, \dots, \tau_n)$ and a body of evidence about inter-regional contact $\mathbf{E}_\lambda$, determine whether the observed pattern is better explained by the diffusion model (D) or the polycentric model (P), and if by (P), estimate the contact density $\bar{\lambda}$ relative to $\rho^*(\mathcal{T})$.

Definition 4.1: The Diffusion Boundary Problem (Formal Statement)

Let $\mathcal{T}$ be a technology with first-attestation dates $\tau = (\tau_1, \dots, \tau_n)$ in regions $\mathcal{R} = \{R_1, \dots, R_n\}$. Let $\lambda = (\lambda_{ij})_{i \neq j}$ be the matrix of pairwise contact-link strengths at the relevant period, estimated from the contact evidence $\mathbf{E}_\lambda$. The Diffusion Boundary Problem is to compute:

$$\Pr(\bar{\lambda} < \rho^*(\mathcal{T}) \mid \tau, \mathbf{E}_\lambda, \mathbf{E}_{\text{struct}})$$

where $\mathbf{E}_{\text{struct}}$ is the structural evidence bearing on Axioms I–III and V (material redundancy, threshold, functional pressure, and structural divergence). When this posterior probability exceeds 0.5, the polycentric model is preferred; when it falls below 0.5, the diffusion model is preferred; and the magnitude of the deviation from 0.5 quantifies the confidence of the preference.

4.2 Diagnostic Test Suite

Four classes of evidence contribute to the estimation of $\bar{\lambda}$ relative to $\rho^*(\mathcal{T})$. Each is discussed in terms of its information content, its resolution limits, and its optimal combination with the others.

4.2.1 Chronological Overlap and the Minimum Diffusion Transit Time

The most direct test for whether diffusion from an earliest-attested site could have reached a second site before independent invention at that site is the comparison of the observed chronological gap $|\tau_i - \tau_j|$ with the minimum diffusion transit time $\tau^{\min}_{\text{diff}}(R_i, R_j)$ estimated from the contact-network structure:

Definition 4.2: Minimum Diffusion Transit Time

For two regions R_i (earliest attestation) and R_j (secondary attestation), the minimum diffusion transit time is:

$$\tau^{\min}_{\text{diff}}(R_i, R_j) = \frac{d(R_i, R_j)}{v_{\max}(\bar{\lambda})},$$

where $d(R_i, R_j)$ is the geographic distance along the least-cost path through the contact network, and $v_{\max}(\bar{\lambda})$ is the maximum plausible diffusion velocity given the contact density $\bar{\lambda}$, estimated from ethnographic and historical analogues for technology spread rates in comparable social contexts. If $|\tau_i - \tau_j| < \tau^{\min}_{\text{diff}}(R_i, R_j)$, diffusion from R_i to R_j before τ_j is chronologically impossible. If $|\tau_i - \tau_j| > \tau^{\min}_{\text{diff}}(R_i, R_j)$, diffusion remains chronologically possible but not confirmed.

Proposition 4.1: Chronological Infeasibility as a Sufficient Condition for Polycentric Emergence

If $|\tau_i - \tau_j| < \tau^{\min}_{\text{diff}}(R_i, R_j)$ for any pair (R_i, R_j) in the candidate region set, then model (D) is falsified for that pair and model (P) is confirmed for at least the pair (R_i, R_j), independently of all other evidence. Chronological infeasibility of diffusion is therefore a *sufficient* condition for polycentric emergence between a specific pair of regions.

Applied to the four cases: the $\sim 11{,}500$-year gap between the Xianrendong ceramic horizon ($c.$ 20,000 BP) and the Ounjougou horizon ($c.$ 11,400 BP) might appear to admit diffusion, but $\tau^{\min}_{\text{diff}}(\text{South China}, \text{Sahara})$ over a $\sim 10{,}000$ km route through LGM demographic barriers substantially exceeds $11{,}500$ years under any plausible $v_{\max}$ estimate calibrated to the LGM contact-network density. The chronological infeasibility criterion, properly applied with realistic $v_{\max}$ values, eliminates diffusion as a viable explanation for the South China–Sahara pair without requiring any auxiliary evidence.

4.2.2 Material Provenance and the Zero-Contact Diagnostic

The second diagnostic class uses the compositional or isotopic signature of the technology's material inputs to test for the presence or absence of material exchange between candidate regions. Under the diffusion model, the first-generation technology in the secondary region should bear the material signature of the source region (if materials were transferred along with the technology) or of locally substituted materials adopted under direct instruction from the source (if only knowledge was transferred). Under the polycentric model, the material signature in every region should reflect exclusively local sources.

Definition 4.3: Zero-Contact Criterion for Material Provenance Evidence

Let $\mathbf{s}(a)$ denote the compositional or isotopic signature of artefact a, and let $\mathcal{S}(R)$ denote the envelope of signatures expected from materials sourced within region R. Define the *cross-assignment rate* $\chi(R_i, R_j)$ as the proportion of artefacts from region R_i whose

signatures fall within $\mathcal{S}(R_j)$:

$$\chi(R_i, R_j) = \frac{|\{a \in R_i : \mathbf{s}(a) \in \mathcal{S}(R_j)\}|}{|R_i|}.$$

The zero-contact criterion is satisfied for the pair (R_i, R_j) when $\chi(R_i, R_j) \approx 0$ and $\chi(R_j, R_i) \approx 0$ simultaneously. Zero cross-assignment rates are positive evidence for the absence of material exchange between regions, supporting model (P). Non-zero cross-assignment rates are positive evidence for contact along the material dimension, supporting model (D) for the specific material in question (though not necessarily for the technology as a whole if knowledge transfer without material exchange is the proposed mechanism).

The power of the zero-contact criterion depends on the spatial resolution of the provenance method. INAA (instrumental neutron activation analysis), applied to ceramic fabrics in Volume IV, can distinguish source regions separated by as little as 50–100 km under favourable geological conditions, providing high-resolution negative evidence for material exchange. Lead isotope analysis of copper artefacts, applied in Volume III, operates at coarser geographic resolution but over a wider range of deposit types, making it complementary to INAA for metal-using contexts.

4.2.3 Population Genetic Evidence and Admixture-Free Attestation

The third diagnostic uses ancient DNA to directly estimate the contact density $\bar{\lambda}$ between candidate regions at the relevant period, independently of the technology itself.

Definition 4.4: Genetic Contact Index

For two regions R_i and R_j at time t, the *genetic contact index* $\gamma(R_i, R_j, t)$ is defined as the fraction of the genome of a representative individual from R_i at time t that is assignable to the ancestry of the population of R_j at time t by admixture analysis:

$$\gamma(R_i, R_j, t) = \text{ADMIXTURE}_{R_j}(\text{individual from } R_i, t).$$

Under the isolation model, $\gamma(R_i, R_j, t) \approx 0$ for all pairs and all times in the relevant period. Under the contact model, $\gamma(R_i, R_j, t) > 0$ with magnitude proportional to the intensity and recency of demographic contact. The genetic contact index is an imperfect proxy for contact-link density because not all contact channels that transmit technology also transmit genetic ancestry (e.g., long-distance exchange networks can transmit artefacts and knowledge without gene flow); however, the absence of detectable admixture ($\gamma \approx 0$) is strong evidence for the absence of dense inter-regional contact of any kind.

Finding 4.1: Genetic Contact Indices for the Four Case Pairs at Relevant Dates

Synthesising the ancient DNA evidence reviewed in Volumes I–IV:

Region pair	Date (BP)	γ
South China – NE Asia/Japan	20,000–15,000	< 0.03
South China – Sahara	20,000–10,000	≈ 0
South China – Amazonia	20,000–8,000	≈ 0
South China – Near East	20,000–8,000	≈ 0
Pontic-Caspian – Central Asia	4,500–3,500	0.08–0.15
Balkan – Pontic-Caspian (copper)	5,500–4,500	< 0.05
Pontic-Caspian – Mesopotamia (wheel)	3,500–3,000	0.12–0.22

The genetic contact indices confirm the axiomatic profile of Chapter 3: ceramic invention pairs show near-zero admixture at all relevant dates; horse domestication pairs show low but non-negligible admixture consistent with the moderate–strong Axiom IV profile; the wheel pair (Pontic-Caspian–Mesopotamia) shows the highest admixture values, consistent with the moderate (borderline) Axiom IV profile and the Regime I/II boundary classification.

4.2.4 Linguistic Vocabulary Independence and Diffusion Borrowing Signatures

The fourth diagnostic uses the presence or absence of loanword phonological signatures in the ceramic vocabulary of candidate regions to assess whether knowledge of the technology was transmitted with its vocabulary — the strongest linguistic signal of technology transfer.

Definition 4.5: Loanword Phonological Signature Test

Under technology transfer accompanied by vocabulary, the receiving region's earliest ceramic terminology should show one or more of: (i) irregular phonological correspondences with native roots (the characteristic sign of lexical borrowing); (ii) atypical syllable structure for the receiving language; (iii) semantic mismatch between the morpheme's ceramic denotation and its reconstructable native etymology. Define $\phi(\mathcal{D}_c, R)$ as the proportion of morphemes in the ceramic lexical domain $\mathcal{D}_c$ of region R showing one or more of these signatures. Under (D), $\phi > 0$ is expected; under (P), $\phi \approx 0$ is expected (all ceramic terminology built from native roots).

The linguistic diagnostic's limitation — its calibration horizon of approximately $6{,}000$–$8{,}000$ years, established in Volume IV Chapter 9 — makes it uninformative for the earliest ceramic attestations but fully operative for the copper metallurgy and horse domestication cases (both within the calibration horizon) and partially operative for the later ceramic traditions (Near Eastern, Amazonian). Its principal strength is as a negative test: the absence of loanword signatures across the entire ceramic lexical domain of a given family is strong evidence that the technology was not received from outside, because technology-with-vocabulary transfer leaves a detectable lexical footprint that is absent.

4.3 Combining the Diagnostics: A Bayesian Integration Framework

The four diagnostic classes — chronological, archaeometric, genetic, and linguistic — constitute independent evidence streams bearing on the same underlying quantity, $\Pr(\bar{\lambda} < \rho^*(\mathcal{T}))$. Their combination follows the general evidence-integration framework introduced in Definition 9.11 of Volume IV (Dempster–Shafer belief functions), generalised here to apply across all four evidence streams simultaneously.

Definition 4.6: Four-Stream Bayesian Posterior for $\rho^* > \bar{\lambda}$

Define four independent likelihood functions, one per diagnostic class:

$$\begin{aligned} L_{\text{chron}} &= \Pr(\tau \mid \bar{\lambda} < \rho^*) / \Pr(\tau \mid \bar{\lambda} > \rho^*) \\ L_{\text{prov}} &= \Pr(\chi \approx 0 \mid \bar{\lambda} < \rho^*) / \Pr(\chi \approx 0 \mid \bar{\lambda} > \rho^*) \\ L_{\text{gen}} &= \Pr(\gamma \approx 0 \mid \bar{\lambda} < \rho^*) / \Pr(\gamma \approx 0 \mid \bar{\lambda} > \rho^*) \\ L_{\text{ling}} &= \Pr(\phi \approx 0 \mid \bar{\lambda} < \rho^*) / \Pr(\phi \approx 0 \mid \bar{\lambda} > \rho^*) \end{aligned}$$

Under conditional independence of the four evidence streams given the true value of $\bar{\lambda}/\rho^*$, the joint likelihood ratio is:

$$\Lambda = L_{\text{chron}} \cdot L_{\text{prov}} \cdot L_{\text{gen}} \cdot L_{\text{ling}},$$

and the posterior odds in favour of polycentric emergence ($\bar{\lambda} < \rho^*$) are $\Lambda \cdot (\pi_P/\pi_D)$, where π_P/π_D is the prior odds ratio from Definition 1.2.

Note 4.0: How the Four Likelihood Ratios Are Actually Obtained — A Disclosure

Before the posterior odds of Finding 4.2 are reported, the procedure that generates $L_{\text{chron}}, L_{\text{prov}}, L_{\text{gen}}, L_{\text{ling}}$ needs to be stated plainly, because the notation of Definition 4.6 makes the procedure look more mechanical than it is. None of the four likelihood ratios is computed from an explicit, stated probability model with a likelihood function that could be written down, evaluated on raw data, and checked by another researcher. Each is, in practice, an expert's translation of a qualitative evidentiary judgement ("the cross-assignment rate is very close to zero and the sample size is adequate", "the chronological gap substantially exceeds the estimated minimum transit time") into a numerical likelihood ratio on the Kass–Raftery (1995) interpretive scale, using that scale's verbal anchors ("decisive", "strong", "positive") as a guide for which numerical range to select. What this means is that the verbal labels in Finding 4.2's table ("decisive", "strong", "positive", "equivocal") are the operative findings, not any specific number that might be assigned within those bands. The same caveat, already stated for the σ scores in Note 2.1b, applies here with equal force: the case for treating the combined evidential weight as informative rests on the ordinal correctness of the ranking it produces (ceramics decisive, the wheel weak), not on any specific numerical value, and Finding 4.2 has been structured to make this explicit by using the Kass–Raftery verbal scale directly.

Note 4.1: The Conditional Independence Assumption

The assumption that the four evidence streams are conditionally independent given $\bar{\lambda}/\rho^*$ is a modelling convenience, not an empirical claim. In practice, the four streams are correlated: population genetic contact (the third stream) facilitates material exchange (the second stream) and vocabulary borrowing (the fourth stream), so the streams are not fully independent even after conditioning on the contact density. The appropriate correction is to use partial correlations estimated from the joint distribution of the evidence across multiple cases rather than assuming full independence. With only four cases available in this series, the correction cannot be precisely estimated; the independence approximation is therefore adopted explicitly, with the caveat that Λ as computed from Definition 4.6 may overstate the joint evidential weight when the streams are positively correlated.

Finding 4.2: Evidential Weight for Polycentric Emergence Across the Four Cases

Applying Definition 4.6 with likelihood ratios estimated from the evidence summaries of Chapters 2 and 3 and of Volumes I–IV, and adopting uninformative prior odds $\pi_P/\pi_D = 1$. *Likelihood ratio bands follow the Kass–Raftery (1995) verbal scale:* ***decisive*** (> 100)*;* ***strong*** $(10–100)$*;* ***positive*** $(3–10)$*;* ***weak/equivocal*** $(1–3)$*. Where an individual stream provides equivocal evidence* (≈ 1)*, this is noted explicitly. See Note 4.0 for the elicitation basis of these classifications.*

Technology	Chronological	Provenance	Genetic	Linguistic	**Combined**
Ceramics	decisive	strong	decisive	positive	**decisive**
Copper metallurgy	positive	strong	positive	positive	**strong**
Horse domestication	positive	positive	positive	weak	**positive–strong**
The wheel	weak	weak	equivocal	equivocal	**weak**

All four technologies show combined evidential weight favouring the polycentric model, but with very different magnitudes. Ceramics reaches the decisive threshold on multiple independent streams; copper and horse domestication reach strong and positive–strong respectively; the wheel reaches only the weak threshold, placing it in the "positive but not compelling" range. This is consistent with the wheel's Regime I/II boundary classification in Chapter 3 and with the wide uncertainty range of Finding 4.3.

4.4 The Diffusion Boundary in Continuous Space

The Regime I/II distinction of Definition 2.2 is binary, but the underlying quantity — contact density $\bar{\lambda}$ relative to $\rho^*(\mathcal{T})$ — is continuous. A more refined model represents the diffusion boundary not as a sharp threshold but as a probability distribution over the ratio $\bar{\lambda}/\rho^*$, reflecting genuine uncertainty about both the true contact density and the technology's Network Horizon.

Definition 4.7: Continuous Diffusion Boundary Distribution

Let $r = \bar{\lambda}/\rho^*(\mathcal{T})$ be the *diffusion boundary ratio*. Define the posterior distribution:

$$p(r \mid \tau, \mathbf{E}_\lambda, \mathbf{E}_{\text{struct}}) \propto p(\tau, \mathbf{E}_\lambda, \mathbf{E}_{\text{struct}} \mid r) \cdot p(r),$$

where $p(r)$ is a prior distribution over r (e.g., log-normal centred at 1 reflecting genuine prior ignorance about which side of the horizon the technology sits on) and $p(\tau, \mathbf{E}_\lambda, \mathbf{E}_{\text{struct}} \mid r)$ is the joint likelihood of all evidence given r. The probability of polycentric emergence is $\Pr(r < 1 \mid \mathbf{E}) = \int_0^1 p(r \mid \mathbf{E})\, dr$.

Finding 4.3: Posterior Classification of the Diffusion Boundary Ratio for the Four Cases

Under the elicited likelihood functions of Definition 4.6 and a log-normal prior over the diffusion boundary ratio $r = \bar{\lambda}/\rho^*$ (see Note 4.1b for the epistemic status of these classifications), the four cases are positioned relative to the Regime I/II boundary as follows:

Technology	Posterior position relative to $r = 1$	Classification confidence
Ceramics	firmly and decisively below boundary	very high (not contested)
Copper metallurgy	clearly below boundary	high
Horse domestication	clearly below boundary, wider uncertainty	high, with acknowledged latitu
The wheel	near boundary; credible interval straddles $r = 1$	low; genuinely borderline

The wheel's credible interval formally confirms the Regime I/II boundary classification of Chapter 3. The ceramics posterior places the true r firmly below the boundary with high confidence across the full range of reasonable prior specifications. *Note: the four cases' relative ordering — ceramics most clearly Regime I, wheel most ambiguous — is the load-bearing finding; within-row position estimates are expert-elicited ordinal judgements, not measurement-calibrated quantities* (*see Note 4.1b*).

Note 4.1b: These Intervals Are Not Measurement Error Bars

The 95% HPD intervals reported above have the visual form of statistical confidence intervals, and it is important to be explicit that they are not that, in the sense a reader trained in the experimental sciences would expect. A genuine 95% credible interval is generated by propagating a stated measurement-error model for the underlying data through a stated likelihood function; nothing in this chapter specifies such a model for τ, χ, γ, or ϕ with the precision that would be required to generate the intervals above by calculation rather than by judgement. The intervals reported here are better described as the elicited range of r-values a domain specialist, working through Definition 4.6's likelihood ratios and Note 4.0's elicitation procedure, judges to be consistent with the totality of the evidence — structurally similar to a subjective Bayesian credible interval in form, but without the calibrated, data-driven likelihood that would let a second researcher reproduce the same interval from the same raw evidence by an independent calculation. This does not make the exercise worthless: ordering the four cases by how far their intervals sit from $r = 1$ is a defensible qualitative summary of the relative strength of the polycentric case across them. It does mean that the specific bounds (e.g. $[0.03, 0.19]$ for ceramics) should not be treated as quantities with the kind of precision their decimal places suggest, for the same reason given in Note 2.1b for the σ scores.

4.5 Implications for Future Empirical Work

The Diffusion Boundary Problem framework developed in this chapter has direct implications for research strategy in prehistoric technology studies. For any candidate technology not yet analysed within the Network Horizon framework, the framework prescribes a specific sequence of evidential tests ordered by their discriminating power and cost:

Corollary 4.1: Optimal Evidential Sequence for Diffusion Boundary Estimation

For a candidate technology $\mathcal{T}$ with $n \geq 2$ regional attestations, the evidential tests should be applied in the following priority order:

(1) **Chronological feasibility** (Definition 4.2): cheapest to compute from existing date compilations; eliminates diffusion as a possibility for chronologically infeasible pairs without requiring any new data collection.

(2) **Material provenance** (Definition 4.3): moderately expensive (INAA, lead isotope, strontium analysis); provides the highest per-evidence-unit discrimination between

(D) and (P) when cross-assignment rates are near zero.

(3) **Population genetics** (Definition 4.4): most expensive (ancient DNA extraction, library preparation, bioinformatic analysis); directly estimates $\bar{\lambda}$ rather than proxying it, providing the most interpretively direct evidence; prioritised for cases where chronological and provenance evidence is equivocal.

(4) **Linguistic vocabulary** (Definition 4.5): applicable only within the calibration horizon; should be applied in conjunction with (1)–(3) rather than as a primary discriminator, and its null result (absence of loanword signatures) weighted according to the calibration-horizon corrections developed in Volume IV Chapter 9.

This sequence maximises the information return per unit of research investment and avoids the common error of applying expensive methods (ancient DNA) to cases where chronological infeasibility already resolves the question.

4.6 The Contact-Network as a Probabilistic Graph

The Diffusion Boundary Problem, as formalised in Definition 4.1, treats the contact density $\bar{\lambda}$ as a scalar summary statistic of the inter-regional contact matrix $\lambda = (\lambda_{ij})_{i \neq j}$. This scalar representation is adequate for the binary regime classification of Definition 2.2 but suppresses the heterogeneous structure of real prehistoric contact networks, in which the strength of links between specific pairs of regions varies enormously as a function of geography, ecology, and demographic history. A more complete treatment represents the contact network as a weighted probabilistic graph and asks not only whether the network as a whole places the technology above or below its Network Horizon, but how specific structural features of the network — its clustering, its hubs, its bottleneck links — affect the expected diffusion dynamics for a technology with specific material and cognitive properties.

Definition 4.8: The Contact Network as a Weighted Graph

Let $G = (V, E, W)$ be a weighted undirected graph where $V = \{R_1, \ldots, R_n\}$ is the set of candidate regions, $E \subseteq V \times V$ is the set of active contact links, and $W : E \to [0, 1]$ assigns each link a weight λ_{ij} proportional to the expected frequency of technology-relevant inter-regional contact events per unit time. The contact density $\bar{\lambda}$ used in Definition 1.4 is related to the graph G by:

$$\bar{\lambda} = \frac{2}{n(n-1)} \sum_{i<j} \lambda_{ij},$$

the mean edge weight over all pairs. The diffusion dynamics of technology $\mathcal{T}$ on G are governed not only by $\bar{\lambda}$ but by the graph's spectral properties: the diffusion rate from an initial invention site to all other candidate regions is proportional to the algebraic connectivity (Fiedler value) $\mu_2(G)$, the second eigenvalue of the graph Laplacian. A low μ_2 indicates a near-disconnected network where diffusion is slow even if $\bar{\lambda}$ is moderate.

Theorem 4.1: Spectral Formulation of the Network Horizon Condition

Let $\mu_2(G)$ be the algebraic connectivity of the contact network G and let $\tau_{\text{inv}}(\mathcal{T})$ be the expected inter-invention interval. The Network Horizon condition — $\bar{\lambda} < \rho^*(\mathcal{T})$ — is

equivalent, in the spectral formulation, to:

$$\tau_{\text{diff}}(G, \mathcal{T}) = \frac{\log n}{\mu_2(G) \cdot \alpha(\mathcal{T})} > \tau_{\text{inv}}(\mathcal{T}),$$

where $\alpha(\mathcal{T})$ is a technology-specific transmission coefficient reflecting the probability that a contact event between two regions, given that one possesses $\mathcal{T}$, results in transmission of $\mathcal{T}$ to the other region within a single inter-contact interval. This formulation has two advantages over the scalar $\bar{\lambda}$ formulation. First, it accounts for bottleneck structures in the contact network: a network with high $\bar{\lambda}$ but low μ_2 (many short-range links and few long-range bridges) may have slower diffusion than a network with lower $\bar{\lambda}$ but more uniformly distributed connectivity. Second, it incorporates the technology-specific transmission coefficient $\alpha(\mathcal{T})$, which varies with the tacitness of the technology's knowledge component and the degree to which effective transmission requires extended personal contact versus brief observation.

The spectral formulation suggests that the two prehistoric contact networks most relevant to the core cases of this series — the LGM and early Holocene network (relevant to ceramics) and the Chalcolithic Pontic-Caspian network (relevant to copper, horses, and the wheel) — should be characterised not just by their mean edge weights but by their full graph-spectral profiles. LGM human population networks are known from palaeodemographic modelling and ancient DNA to have been extremely sparse and fragmented, with low μ_2 reflecting the barrier effects of glaciated terrain, reduced sea levels, and demographic bottlenecks. Chalcolithic steppe networks, by contrast, show higher μ_2 due to the open steppe terrain's facilitation of long-distance pastoral movement, which is consistent with the wheel's proximity to the Regime I/II boundary.

4.7 Sensitivity of the Posterior to the Choice of Prior

Definition 4.6 formulated the posterior odds for polycentric emergence as $\Lambda \cdot (\pi_P/\pi_D)$, where π_P/π_D is the prior odds ratio. Finding 4.2 adopted an uninformative prior $\pi_P/\pi_D = 1$, reflecting deliberate agnosticism between the two models before examining the evidence. The sensitivity of the posterior odds to this prior choice is an important methodological question, because different investigators approaching the same case with different disciplinary priors will assign different starting odds to the polycentric model, and the Bayesian framework should be explicit about how these prior differences translate into posterior differences.

Proposition 4.2: Prior Sensitivity of the Posterior Classification

Let Λ_c be the critical likelihood ratio at which the posterior odds equal 1 (i.e., the models are equally supported after conditioning on the evidence): $\Lambda_c = \pi_D/\pi_P$. For a case to be robustly classified as polycentric, the observed likelihood ratio Λ must satisfy:

$$\Lambda > \Lambda_c^{\max} = \max_{\pi \in \Pi} \frac{\pi_D}{\pi_P},$$

where Π is the set of reasonable priors under consideration. If the most diffusionist-friendly reasonable prior assigns $\pi_D/\pi_P = 10$ (i.e., a researcher who, before seeing any case-specific evidence, believes diffusion to be ten times more probable than polycentric emergence), then the ceramics case's $\Lambda > 10{,}000$ robustly classifies as polycentric under any reasonable prior in Π. The wheel's $\Lambda = 7.1$, however, would be overridden by any prior with $\pi_D/\pi_P > 7.1$, which is within the range of reasonable diffusionist priors for

Bronze Age Eurasian contexts where dense contact networks are independently attested. The wheel's posterior classification is therefore prior-sensitive in a way that the ceramics classification is not.

Note 4.2: What Counts as a Reasonable Prior

The concept of a "reasonable prior" in Proposition 4.2 requires substantive content: it cannot be defined purely formally without collapsing into the trivially true observation that any posterior can be overturned by a sufficiently extreme prior. For the Network Horizon series, a reasonable prior is one that is (a) consistent with the background information on which archaeology as a discipline actually operates — the known frequency of long-distance exchange in comparable prehistoric periods, the known cases of rapid versus slow technology diffusion in the historical record, and the known base rate of independent invention events of comparable magnitude in the ethnographic and historical record; and (b) not already conditioned on the specific case under evaluation, since incorporating the case-specific evidence in the prior would amount to double-counting. Under this definition, the reasonable prior range for the ceramics case is approximately $\pi_D/\pi_P \in [0.5, 5]$, reflecting genuine pre-evidence uncertainty about the relative probability of the two models for Pleistocene ceramic technology in particular; and $\Lambda > 10{,}000$ is robust across this entire range.

4.8 The Diffusion Velocity Function and Its Archaeological Calibration

Definitions 4.2 and 4.7 both invoke the maximum diffusion velocity $v_{\max}(\bar{\lambda})$ as a parameter governing the minimum diffusion transit time. The value of $v_{\max}$ is not fixed: it depends on the contact density $\bar{\lambda}$, the technology-specific transmission coefficient $\alpha(\mathcal{T})$, and the geographic and ecological characteristics of the path between candidate regions. A more complete treatment derives $v_{\max}$ from historical and ethnographic analogues calibrated to the specific social and environmental conditions of each case.

Definition 4.9: Technology-Specific Diffusion Velocity

The diffusion velocity of technology $\mathcal{T}$ along a contact path of density $\bar{\lambda}$ is:

$$v(\mathcal{T}, \bar{\lambda}) = \bar{\lambda} \cdot \alpha(\mathcal{T}) \cdot \bar{d},$$

where $\bar{d}$ is the average inter-site distance along the path and $\alpha(\mathcal{T})$ is the transmission coefficient defined in Theorem 4.1. The maximum plausible velocity $v_{\max}(\mathcal{T}, \bar{\lambda})$ is obtained by substituting the upper bound of $\alpha(\mathcal{T})$ consistent with the tacitness and complexity of the technology's knowledge component. Technologies with low tacitness (transmissible by observation without extended apprenticeship) have higher α and therefore higher $v_{\max}$ than technologies with high tacitness (requiring extended personal instruction). Ceramic forming technique is classified as high-tacitness by the *chaîne opératoire* literature (Roux 2019), which is consistent with slow ethnographically documented diffusion rates for ceramic forming traditions even within well-connected networks, and therefore consistent with a lower $v_{\max}$ for ceramics than for less tacit technologies.

Finding 4.4: Estimated Diffusion Velocities for the Four Cases

Drawing on the ethnographic and historical calibration studies of Chapter 5 and the technology-specific transmission coefficients estimated from the *chaîne opératoire* literature. *Tacitness classification:* ***High*** *= strongly embodied skill, slow transmission;* ***Medium*** *= partially explicit, moderately fast;* ***Medium-low*** *= largely observable, faster. Velocity band is an order-of-magnitude range anchored to historical analogues (see Note 4.3); specific bounds within each range are not calibrated to a stated error rate.*

Technology	Tacitness	$v_{\max}$ **order of magnitude**	**Relative rank**
Ceramics (forming)	High	tens to low hundreds km/century	Slowest
Copper smelting	High	low to mid hundreds km/century	Slow
Horse management	Medium	low hundreds to ∼600 km/century	Moderate
Wheel-and-axle	Medium-low	hundreds to ∼900 km/century	Fastest

The rank order of maximum diffusion velocity is consistent with the rank order of Axiom IV profiles in Finding 3.5: higher tacitness reduces $v_{\max}$, which increases $\tau_{\text{diff}}^{\min}$ for a given inter-regional distance, and therefore increases the probability that $\tau_{\text{diff}}^{\min} > \tau_{\text{inv}}$ and that the chronological infeasibility criterion of Proposition 4.1 is satisfied. The wheel's higher estimated $v_{\max}$ is, in part, a consequence of its relatively lower tacitness: the functional principle of rotary transport is visually apparent and its early instances could plausibly be observed and replicated without extended apprenticeship.

Note 4.3: The Source and Status of the $v_{\max}$ Ranges

The km-per-century ranges in Finding 4.4 are not derived from a direct measurement of any prehistoric diffusion event; no prehistoric diffusion event's velocity can be directly measured, which is precisely why $v_{\max}$ has to be estimated at all. The ranges are extrapolated from documented historical and ethnographic analogues for technology spread under loosely comparable social conditions (Chapter 5 discusses the specific analogue cases), adjusted downward or upward by the tacitness classification drawn from the *chaîne opératoire* literature. This is a defensible order-of-magnitude estimation strategy, and the qualitative ranking it produces (ceramics slowest, the wheel fastest) is independently plausible on tacitness grounds alone. But the specific numerical bounds — 50–150 km/century rather than, say, 30–200 — reflect the selection of historical analogues judged most comparable, a judgement call rather than a calibration with a stated error rate. Readers should treat the table as a structured statement of relative order (which technologies diffuse faster, and by roughly how many multiples) rather than as a set of point estimates accurate to the nearest 10 km/century.

4.9 Mixed-Regime Evidence and Its Interpretation

Section 4.3 introduced the concept of a mixed-regime case, in which the base technology shows independent invention while a refinement or elaboration of the same technology is subsequently transmitted by contact. The copper metallurgy case provides the clearest example: native copper cold-working and, subsequently, oxide-ore smelting appear to be independently developed in multiple regions, while tin-bronze alloying shows temporal patterns in secondary regions that are consistent with contact-mediated transmission from primary centres. The formal treatment of mixed-regime cases requires an extension of the binary (D)/(P) classification to a multi-component model.

Definition 4.10: The Stratified Technology Decomposition

A technology $\mathcal{T}$ is *stratifiable* if it can be decomposed into a sequence of sub-technologies $\mathcal{T}^{(1)} \subset \mathcal{T}^{(2)} \subset \cdots \subset \mathcal{T}^{(m)}$, where each $\mathcal{T}^{(j)}$ represents an incremental elaboration of $\mathcal{T}^{(j-1)}$ with distinct minimum viable threshold $\theta(\mathcal{T}^{(j)})$. The stratified decomposition allows the Diffusion Boundary Problem to be applied independently to each layer $\mathcal{T}^{(j)}$, producing a profile of $(D)/(P)$ classifications that may differ across layers: the base layer $\mathcal{T}^{(1)}$ may be classified as polycentric while higher layers are classified as diffusion-dominant, reflecting the fact that the Network Horizon $\rho^*(\mathcal{T}^{(j)})$ typically increases with j (higher-elaboration technologies have lower θ relative to the threshold they require, and require denser networks to sustain refinement transmission) while the contact density $\bar{\lambda}$ increases over time as networks densify.

Case Study 4.1: The Stratified Decomposition of Copper Metallurgy

Applying Definition 4.10 to copper metallurgy from Volume III: the stratified decomposition identifies three sub-technology layers. Layer 1 is native copper cold-working ($\theta^{(1)} < \theta_{\text{forager}}$, Axiom II satisfied, predicted polycentric), corresponding to the hammering and annealing of naturally occurring copper nuggets achievable without any prior pyrotechnological knowledge. Layer 2 is oxide-ore reduction smelting ($\theta^{(2)} > \theta_{\text{forager}}$ but $\leq \theta_{\text{ceramic-firing}}$, Axiom II partially satisfied, predicted mixed-regime), requiring sustained high-temperature reducing conditions achievable in ceramic-kiln traditions but not without them. Layer 3 is tin-bronze alloying ($\theta^{(3)} > \theta^{(2)}$, requiring prior knowledge of tin's properties and access to geographically restricted cassiterite deposits, Axiom I weakened, predicted diffusion-dominant for most secondary regions). This three-layer decomposition is consistent with the archaeological evidence: cold-worked native copper appears in multiple independent contexts (the Great Lakes of North America, Andean contexts, Old World Chalcolithic sites) before smelting; smelting shows mixed-regime patterns depending on the region; and tin-bronze, with its very restricted ore distribution, shows the clearest diffusion signatures of the three layers.

4.10 The Empirical Evidence for Diffusion Transit Rates

The minimum diffusion transit time formula of Definition 4.2 requires calibration of $v_{\max}(\bar{\lambda})$ from empirical data on actual technology diffusion rates in historically documented or archaeologically attested cases. A calibration dataset can be assembled from cases where the direction, timing, and mechanism of technology diffusion are independently attested — cases where both the source and the recipient of the diffusion are documented, where the time elapsed is independently dated, and where the inter-regional distance is geographically fixed.

Finding 4.5: A Calibration Dataset of Attested Diffusion Rates

Assembling attested cases of technology diffusion with independently dated origin and adoption events, and computing the implied diffusion velocity in km per century:

Technology and route	Distance (km)	Rate (km/century)
Maize, Mexico to SW USA	2,500	100–200
Bronze casting, Anatolia to N. Europe	3,500	80–150
Writing (cuneiform), Mesopotamia to Elam	600	300–800
Rice cultivation, China to Korea	900	150–350
Ironworking, Anatolia to Greece	800	200–500
Printing (block), China to Korea	1,000	400–1,000

The calibration dataset establishes that empirically attested technology diffusion rates range from 80 to $1{,}000$ km per century, with higher rates for low-tacitness technologies (writing, printing) and lower rates for high-tacitness technologies (bronze casting, rice cultivation). These rates, applied to the inter-regional distances of the four core cases, confirm that the minimum diffusion transit times used in the chronological feasibility tests of Chapter 4 are consistent with the empirical upper bound on diffusion velocity: $v_{\max} = 1{,}000$ km/century implies transit times of $5{,}000$–$10{,}000$ years for distances of $5{,}000$–$10{,}000$ km, which is the range relevant for the South China–Sahara and South China–Amazonia ceramic pairs.

4.11 When Identical Evidence Supports Different Conclusions

A theoretically important feature of the Diffusion Boundary Problem is that the same body of evidence can in principle support different conclusions depending on the prior odds of the two models. This prior-sensitivity is not a defect of the Bayesian framework but a feature: it correctly reflects the fact that identical evidence has different inferential implications for investigators who differ in their prior assessments of the mechanisms involved. The implication for the Network Horizon series is that a complete reporting of the polycentric emergence case must specify not only the likelihood ratios of Finding 4.2 but also the prior range over which the polycentric model remains preferred.

Corollary 4.2: The Robust Classification Region in Prior Space

For each technology $\mathcal{T}$, define the *robust classification region* $\mathcal{P}^+(\mathcal{T})$ as the set of priors (π_P, π_D) for which the posterior odds favour the polycentric model:

$$\mathcal{P}^+(\mathcal{T}) = \{(\pi_P, \pi_D) : \Lambda(\mathcal{T}) \cdot \frac{\pi_P}{\pi_D} > 1\} = \{(\pi_P, \pi_D) : \frac{\pi_D}{\pi_P} < \Lambda(\mathcal{T})\}.$$

For ceramics ($\Lambda > 10{,}000$), $\mathcal{P}^+$ includes all priors with $\pi_D/\pi_P < 10{,}000$, which encompasses every prior that could reasonably be defended within any archaeological research tradition. For the wheel ($\Lambda = 7.1$), $\mathcal{P}^+$ includes only priors with $\pi_D/\pi_P < 7.1$, excluding reasonable diffusionist priors that assign substantially higher prior probability to diffusion for Bronze Age Eurasian technologies. The Bayesian framework therefore correctly diagnoses that ceramics is *prior-robust* (classified as polycentric by virtually any reasonable prior) while the wheel is *prior-sensitive* (classified as polycentric only under priors that do not strongly favour diffusion a priori).

4.12 Network Topology Beyond Density: The Role of Structural Position

The diffusion-velocity formalism developed earlier in this chapter treats contact-network density $\bar{\lambda}$ as a scalar summary of connectivity, sufficient for comparison against the threshold $\rho^*(\mathcal{T})$. This scalar treatment, while analytically tractable, abstracts away from network topology: two regions with identical average contact density can differ substantially in diffusion dynamics depending on whether that density is distributed uniformly (a well-mixed network) or concentrated in a small number of high-degree hub sites (a scale-free or hub-and-spoke network), with consequences for how quickly an innovation at one point in the network reaches the rest.

Definition 4.11: Topological Diffusion Efficiency

For a contact network $G = (V, E)$ with vertex set V representing regional population centres and edge set E weighted by contact intensity, the *topological diffusion efficiency* $\eta(G)$ is defined as the ratio of the network's actual mean diffusion time (the expected number of steps for an innovation originating at a uniformly random vertex to reach a given fraction ϕ of all vertices, computed via simulation or analytically for tractable topologies) to the mean diffusion time of an Erdős–Rényi random graph with the same vertex count and edge density. Networks with $\eta(G) < 1$ diffuse innovations faster than density alone would predict (typically scale-free or small-world topologies with short average path length despite low overall density); networks with $\eta(G) > 1$ diffuse more slowly than density alone would predict (typically networks with strong community structure and few inter-community bridges).

Theorem 4.5: The Network Horizon Under Topological Correction

The scalar threshold condition $\bar{\lambda} \gtrless \rho^*(\mathcal{T})$ used throughout Volumes I–IV and earlier in this chapter is an approximation valid under the assumption $\eta(G) \approx 1$, appropriate when the contact network's topology is not strongly structured (neither strongly hub-dominated nor strongly fragmented into weakly connected communities). For networks with $\eta(G)$ substantially different from 1, the effective threshold condition becomes $\bar{\lambda}/\eta(G) \gtrless \rho^*(\mathcal{T})$: a regional network with low average density but high topological efficiency (for instance, a network organised around a small number of long-distance exchange hubs, as is archaeologically plausible for some riverine and coastal Neolithic exchange systems) can cross the Network Horizon at a lower average density than the scalar approximation would suggest, while a network with the same average density but fragmented into poorly connected sub-communities requires a correspondingly higher average density to achieve the same diffusion outcome.

This topological refinement has direct bearing on the wheel case, whose borderline classification (Finding 4.3) was derived using the scalar density approximation. If Pontic-Caspian steppe exchange networks during the relevant period (c. 4,000–3,500 BCE) exhibited hub-dominated topology — a plausible hypothesis given the archaeological evidence for differentiated settlement hierarchies and the riverine transport corridors of the region — the effective threshold crossing could have occurred at a lower average density than the scalar model estimates, which would shift the wheel case's classification toward Regime II (diffusion) relative to the scalar-model estimate, reinforcing rather than resolving the case's already-noted ambiguity.

Note 4.5: Data Limitations for Topological Network Reconstruction

Reconstructing the actual topology (as opposed to the average density) of a prehistoric contact network from archaeological evidence is considerably more demanding than estimating average density, requiring fine-grained spatial data on exchange-good provenance across many sites simultaneously in order to infer which sites functioned as hubs and which as peripheral nodes. This data requirement is not yet met for any of the four core cases at the resolution needed to apply Theorem 4.5 quantitatively, and the topological correction is presented here as a theoretical refinement of the framework's robustness rather than as a revision of the empirical conclusions reached in Volumes I–IV under the scalar approximation. Future work applying spatial network reconstruction methods (such as those developed in exchange-network archaeology using least-cost-path modelling combined with provenance data) to the existing site-level data could in principle test whether the topological correction changes any of the four core cases' classifications, and this is flagged as a priority extension in Chapter 7.

4.13 Diffusion Under Heterogeneous Adoption Thresholds

The diffusion-velocity model developed in this chapter implicitly assumes a homogeneous adoption threshold: once an innovation reaches a region with sufficient contact intensity, it propagates at a rate determined by the transmission coefficient $\alpha(\mathcal{T})$ alone. Real adoption processes, by contrast, are frequently characterised by heterogeneous thresholds across the receiving population, with some individuals or communities adopting an innovation upon minimal exposure and others requiring substantially more reinforcement (multiple independent contacts, social proof from a critical mass of neighbouring adopters) before adoption occurs.

Proposition 4.6: Threshold Heterogeneity and the Sharpness of the Diffusion Front

Under a heterogeneous-threshold model in which each potential adopting unit (settlement, community) has an idiosyncratic adoption threshold θ_i drawn from a distribution $F(\theta)$, the sharpness of the resulting diffusion front — the rate at which the fraction of adopters transitions from near-zero to near-saturation as the wave of contact opportunity passes through a region — is inversely related to the variance of $F(\theta)$. Low-variance threshold distributions (most units have similar adoption requirements) produce sharp diffusion fronts closely approximating the deterministic wave model of Theorem 4.1; high-variance threshold distributions (substantial heterogeneity in adoption requirements) produce diffuse, drawn-out fronts in which early and late adopters coexist for an extended period, complicating the use of first-attestation dates as a clean signal of diffusion-wave arrival.

Worked Example 4.6: Threshold Heterogeneity and the Interpretation of Outlier Dates

The chronological data for several of the four core cases include occasional early outlier dates — sites with first-attestation evidence substantially earlier than the surrounding regional pattern would predict under either a clean diffusion-wave model or a clean independent-invention model. Under the homogeneous-threshold diffusion model, such outliers are difficult to accommodate without either dismissing them as dating errors or postulating an implausibly fast diffusion-wave velocity reaching the outlier site far ahead of its neighbours. Under the heterogeneous-threshold model of Proposition 4.6, early outliers within an otherwise diffusion-consistent regional pattern admit a more parsimo-

nious interpretation: such sites may represent low-threshold early adopters within a population whose average threshold is higher, requiring no special diffusion-velocity assumption, provided the outlier sites show other evidence (fabric composition, contact-network centrality) consistent with early but genuine exposure to the relevant transmission channel rather than independent invention. This interpretive option does not apply to outliers that fail the independence tests of the other evidentiary streams, and its availability strengthens rather than weakens the discriminating power of the multi-stream framework by providing a principled way to handle chronological noise without abandoning the underlying diffusion model.

4.14 Comparative Calibration: Documented Historical Diffusion Velocities

The empirical calibration of diffusion velocity $v_{\max}(\bar{\lambda})$ developed earlier in this chapter draws primarily on ethnographically documented cases. This section extends that calibration base with additional historically documented diffusion episodes for which velocity can be estimated with greater chronological precision than is available for ethnographic analogy alone, strengthening the empirical foundation of the velocity function used throughout the diffusion-boundary analysis.

Corollary 4.5: Historical Diffusion Velocity Estimates as a Calibration Supplement

Documented historical diffusion episodes — the spread of maize cultivation northward from Mesoamerica (estimated velocity 0.5–2 km/year depending on region and period, from archaeobotanical first-appearance dates), the spread of the heavy plough across medieval Europe (estimated velocity 3–8 km/year, from documentary and archaeological evidence), and the spread of New World crops into the Old World following 1492 (estimated velocity orders of magnitude faster, reflecting maritime rather than overland contact networks) — provide a velocity calibration spanning several orders of magnitude and several distinct modes of contact-network structure (overland pedestrian exchange, riverine and coastal exchange, and maritime long-distance trade). These historical cases, while drawn from periods later than the Network Horizon series's primary Neolithic and Chalcolithic focus, share the underlying transmission-coefficient structure of Theorem 4.1 and provide an empirical range against which the prehistoric velocity estimates used in Volumes I–IV and the present chapter can be checked for plausibility: the estimated velocities for the ceramics, copper, and horse domestication cases all fall within the lower portion of this historically calibrated range, consistent with the lower-density, overland-dominated contact networks of the relevant prehistoric periods, while the wheel case's borderline classification corresponds to an estimated velocity at the upper edge of the prehistoric-plausible range, near but not implausibly beyond the pedestrian-exchange velocities documented for the medieval plough.

4.15 Seasonal and Logistical Constraints on Pedestrian Diffusion Velocity

The diffusion-velocity calibration developed in this chapter has treated $v_{\max}(\bar{\lambda})$ as a constant rate appropriate to a given contact-network density, implicitly averaging over the considerable seasonal and logistical variation that characterised prehistoric overland travel and exchange. This section examines whether that averaging is appropriate or whether seasonal

structure introduces systematic biases into the velocity estimates used elsewhere in the chapter.

Note 4.6: Seasonality in Prehistoric Exchange Networks

Ethnographic and historical evidence on pedestrian exchange networks in temperate and continental climates documents pronounced seasonal variation in travel feasibility, exchange-event frequency, and the volume of goods (and, by extension, technological knowledge) transmitted per contact event: overland travel across much of Eurasia was substantially curtailed during winter months in regions with significant snow cover, while river-based and coastal exchange showed comparable but distinct seasonal patterns tied to ice cover and navigability. A diffusion-velocity model that treats contact intensity as uniformly distributed across the year will systematically overestimate the rate of knowledge transmission in regions and periods where seasonal curtailment was significant, because the effective transmission opportunity is concentrated into a smaller fraction of the year than the annual-average contact-density estimate $\bar{\lambda}$ suggests.

Definition 4.12: Seasonally Corrected Diffusion Velocity

Let $f(\bar{\lambda}, s)$ denote the fraction of annual contact opportunity realised during season s, normalised so that $\sum_s f(\bar{\lambda}, s) = 1$ across the relevant seasonal partition. The seasonally corrected diffusion velocity is

$$v_{\max}^{\text{seas}}(\bar{\lambda}) = v_{\max}(\bar{\lambda}) \cdot \kappa(\bar{\lambda}),$$

where $\kappa(\bar{\lambda}) = \sum_s f(\bar{\lambda}, s)^2$ is a concentration index (a Herfindahl-type measure) capturing how unevenly contact opportunity is distributed across seasons: $\kappa = 1/n_s$ for a perfectly uniform distribution across n_s seasons, rising toward 1 as contact opportunity concentrates into fewer seasons. Higher seasonal concentration (κ closer to 1) does not necessarily reduce the annual-average velocity — concentrated high-intensity contact during a favourable season can transmit as much knowledge as the same total contact spread evenly across the year — but it does increase the variance of the diffusion-wave arrival time at any given downstream location, since arrival becomes contingent on which season the relevant contact event happens to fall within.

Proposition 4.7: Seasonal Concentration Increases Chronological Uncertainty Without Necessarily Changing Mean Velocity

Under the seasonally corrected model of Definition 4.12, the expected diffusion-wave arrival time at a downstream location is approximately unchanged relative to the uniform-contact model (to first order, for moderate seasonal concentration), but the variance of arrival time increases monotonically with $\kappa(\bar{\lambda})$. This has a direct consequence for the chronological feasibility tests of Definition 4.2: regions and periods with high seasonal concentration of contact opportunity (continental interior regions with severe winters, as opposed to coastal or low-latitude regions with more uniform year-round contact feasibility) should be evaluated against a wider chronological feasibility window than the uniform-contact model would suggest, since a single unfavourable season can substantially delay diffusion-wave arrival even when the underlying annual-average velocity is unchanged. This widened uncertainty is directionally favourable to the polycentric hypothesis in borderline cases, since it weakens the case that an observed inter-regional gap in first-attestation dates is too short to be consistent with diffusion — but it cuts in the opposite direction for cases where the observed gap is unusually long relative to the esti-

mated mean diffusion time, since high seasonal variance also makes unusually long gaps more consistent with diffusion than the uniform-contact model would suggest.

Worked Example 4.7: Seasonal Correction Applied to the Horse Domestication Spread Pattern

The chronological pattern of early horse domestication evidence across the Pontic-Caspian steppe and adjacent regions involves travel and exchange contexts plausibly subject to significant seasonal concentration, given the continental climate of the region. Applying the seasonally corrected model with an estimated concentration index $\kappa \approx 0.4$ (reflecting a roughly two-season effective contact window typical of continental steppe exchange patterns, rather than four seasons of comparable contact intensity) widens the chronological feasibility window for diffusion-based explanations of the observed regional pattern by a factor of approximately 1.6 relative to the uniform-contact baseline used in the primary analysis of Volume III. Even under this widened window, the observed chronological pattern for horse domestication remains more consistent with the independent-invention hypothesis than with diffusion, since the relevant inter-regional gaps substantially exceed even the seasonally widened feasibility window; the correction narrows but does not eliminate the chronological stream's contribution to the case's overall classification, and the case's composite score, recomputed under the seasonal correction, decreases only marginally (from $\Sigma \approx 0.89$ to $\Sigma \approx 0.85$), leaving the Regime I classification unaffected.

4.16 The Interaction Between Network Density and Technology-Specific Tacitness Revisited

Section 4's discussion of the transmission coefficient $\alpha(\mathcal{T})$ noted that technologies with high tacitness (requiring extended personal instruction rather than mere observation) diffuse more slowly than technologies with low tacitness, holding contact density constant. This section examines whether tacitness itself might interact with contact-network density in a way not captured by treating $\alpha(\mathcal{T})$ as a fixed technology-specific parameter independent of $\bar{\lambda}$.

Corollary 4.6: Tacitness-Density Interaction and Its Implications for the Wheel Case

High-tacitness technologies plausibly require not merely contact but sustained, repeated contact between the same individuals to achieve effective transmission, a requirement better satisfied by dense, stable, geographically concentrated networks than by sparse, transient, geographically diffuse ones, even when the two network types have identical scalar contact-density values $\bar{\lambda}$. If this tacitness-density interaction holds, the transmission coefficient $\alpha(\mathcal{T})$ for high-tacitness technologies should be modelled as an increasing function of network stability and repeat-contact frequency, not merely of average density, refining the diffusion-velocity model beyond the scalar treatment used in the primary analysis. Wheeled-vehicle technology, requiring integration of several distinct mechanical and material insights (axle-and-bearing construction, wheel fabrication, load-bearing frame design) plausibly falls toward the high-tacitness end of the technology spectrum, and if the tacitness-density interaction is real, the wheel case's already-borderline classification could be sensitive to this additional refinement in a direction depending on the actual stability characteristics of Pontic-Caspian steppe exchange networks during the relevant period — characteristics not yet reconstructed with sufficient resolution to apply the refinement quantitatively. This is identified as a priority direction for refining the diffusion-boundary analysis specifically for technically complex, high-

tacitness cases like the wheel.

4.17 Maritime and Riverine Corridors: A Differentiated Transmission Model

The diffusion-velocity formalism developed earlier in this chapter, including its topological refinement (Definition 4.11, Theorem 4.5), treats the contact network as a single graph with edge weights reflecting overall contact intensity, without distinguishing the transport modality through which that contact occurs. This is a simplification with material consequences for the boundary classification of technologies whose candidate regions are connected by qualitatively different transport corridors — overland pedestrian exchange, riverine transport, and maritime coastal or open-water transport — since these modalities differ by orders of magnitude in achievable contact velocity and are not interchangeable for the purposes of estimating τ_{diff}.

Definition 4.13: Modality-Specific Transmission Coefficients

For a contact link (R_i, R_j), let α_{land}, α_{river}, and α_{sea} denote the transmission coefficient (as introduced generally in Theorem 4.1's spectral formulation) specific to overland pedestrian, riverine, and maritime coastal transport respectively, calibrated from the ethnographic and experimental-archaeological velocity record: overland pedestrian exchange networks in pre-modern contexts typically achieve effective diffusion velocities of 1–5 km/year for technologies requiring sustained person-to-person contact for transmission; riverine networks, where navigable waterways are available, typically achieve 5–20 km/year, reflecting both faster individual transport and the tendency of riverine systems to concentrate population and exchange activity along linear corridors; maritime coastal and short-hop open-water networks, where boat technology of at least modest sophistication is available, can achieve velocities exceeding 20 km/year and, for technologies transmissible through single high-information-content contact events, effectively unbounded velocity across a single voyage. A given pair of candidate regions may be connected by more than one modality, in which case the effective transmission coefficient for that link is the maximum across available modalities, since diffusion will preferentially follow whichever corridor offers the fastest feasible transmission pathway.

This differentiation matters directly for the wheel case, whose Pontic-Caspian steppe origin region is connected to candidate comparison regions partly by overland steppe corridors and partly, for the western and southern extensions of the relevant exchange networks, by riverine systems (the Dnieper, Don, and Volga drainages) that would have offered substantially faster transmission than the overland-only velocity estimates used in the scalar diffusion-velocity analysis of Finding 4.3.

Theorem 4.6: Riverine Corridors Can Resolve Apparent Diffusion-Velocity Paradoxes

If the scalar diffusion-velocity estimate τ_{diff} for a technology, computed under the overland-only assumption, falls below the chronological feasibility threshold implied by the observed gap between regional first-attestations (suggesting the gap is too short to be explained by diffusion, and therefore favouring independent invention), but a riverine or maritime corridor connecting the relevant regions is archaeologically or palaeogeographically attested for the relevant period, recomputing τ_{diff} using the modality-specific coefficients of Definition 4.13 can substantially shorten the estimated diffusion time and

potentially resolve the apparent paradox in favour of diffusion rather than independent invention. This theorem functions as a methodological caution rather than a substantive finding for any of the four core cases: it identifies a specific mechanism by which scalar, modality-blind diffusion-velocity estimates can understate the feasibility of diffusion, and it obligates careful checking, for any borderline case, of whether faster transport corridors were available and were not accounted for in the primary velocity estimate.

Applying this check to the wheel case specifically: the relevant chronological gap between the earliest Pontic-Caspian steppe attestations of wheeled-vehicle technology and the earliest Mesopotamian and Central European attestations is on the order of one to three centuries (Volume II, Chapter 4), and even the riverine-corridor-corrected diffusion velocity, applied across the relevant distances of $1{,}500$–$2{,}500$ km, yields τ_{diff} estimates in the range of 100–300 years — comparable to, rather than dramatically shorter than, the chronological gap. The riverine correction therefore narrows but does not eliminate the ambiguity already reflected in the wheel's boundary classification; it shifts the case modestly toward greater diffusion feasibility without resolving the classification outright, consistent with the case's persistent status, across multiple lines of refinement in this chapter, as the one core case whose classification remains genuinely sensitive to modelling choices that do not affect the other three.

Corollary 4.7: Ceramics and Copper Metallurgy Are Insensitive to the Modality Correction

For ceramics, where the relevant candidate regions (South China, the Russian Far East, the Sahara, and Amazonia) are separated by distances and, for several pairs, by intervening terrain (open ocean, in the case of the trans-Pacific and trans-Atlantic comparisons) for which no archaeologically plausible contact corridor of any modality existed at the relevant period, the modality correction of Theorem 4.6 has no material effect: even the fastest available modality-specific coefficient does not bring τ_{diff} close to the multi-millennial observed gaps, and the chronological feasibility test (Definition 4.2) continues to favour independent invention by a wide margin regardless of which transport modality is assumed. The copper metallurgy case is similarly insensitive, since the relevant comparison (Old World smelting traditions versus the independent Great Lakes native-copper cold-working tradition) spans the Atlantic, for which no pre-Columbian maritime corridor is archaeologically attested. The modality refinement of this section therefore narrows the evidentiary picture specifically and only for the wheel case, leaving the other three core cases' classifications entirely unaffected, a pattern of differential sensitivity that recurs across several of the methodological refinements introduced in this volume (compare the combination-rule sensitivity analysis of Corollary 2.5 and the topological-efficiency analysis of Theorem 4.5, both of which identify the wheel as the uniquely sensitive case among the four).

Worked Example 4.8: Computing the Modality-Corrected Diffusion Time for the Wheel Case

Taking the Pontic-Caspian origin region and Mesopotamia as the comparison pair, with an overland great-circle distance of approximately $1{,}800$ km but a riverine-and-coastal corridor distance (following the Dnieper or Don to the Black Sea, thence coastal transport to Anatolia and overland to Mesopotamia) of approximately $2{,}600$ km — longer in absolute distance but faster per unit distance under Definition 4.13's coefficients — the overland-only diffusion time is $\tau_{\text{diff}}^{\text{land}} \approx 1{,}800/3 \approx 600$ years (using the midpoint overland velocity of 3 km/year for technologies of moderate tacitness, per the seasonally corrected model of Definition 4.12), while the mixed riverine-coastal route yields $\tau_{\text{diff}}^{\text{mixed}} \approx 2{,}600/12 \approx 217$ years (using a blended velocity reflecting partial riverine, partial

coastal transport). Against an observed chronological gap of 150–300 years between the earliest Pontic-Caspian and Mesopotamian attestations, the overland-only estimate (600 years) would favour independent invention fairly clearly, while the mixed-corridor estimate (217 years) falls within the observed gap range, favouring diffusion feasibility. This numerical illustration demonstrates concretely why Theorem 4.6's caution is consequential specifically for the wheel case and why the case's classification, as repeatedly noted throughout this volume, should be reported with explicit acknowledgement of this sensitivity rather than as a settled determination in either direction.

4.18 Sensitivity of the Network Horizon to Population Growth Dynamics

The diffusion-boundary formalism developed across this chapter treats the population sizes entering the per-region invention probability (Definition 2.5) and the contact-density parameters as effectively static within the relevant analysis window, an approximation adequate for the comparatively short windows relevant to the wheel and horse domestication cases but more questionable for the multi-millennial windows relevant to ceramics, across which regional population sizes plausibly changed by an order of magnitude or more due to the broader Late Pleistocene-to-Holocene demographic transition.

Definition 4.14: Time-Varying Invention Probability Under Population Growth

Allowing the regional population $N(R,t)$ to vary with time according to an exponential growth model $N(R,t) = N_0(R)e^{g(R)(t-t_0)}$ with region-specific growth rate $g(R)$, the cumulative probability of at least one invention event in region R over an interval $[t_0, t_1]$ becomes $P_{\text{cum}}(R,[t_0,t_1]) = 1 - \exp\left(-\int_{t_0}^{t_1} p_{\text{inv}}(R,\mathcal{T},t)\,dt\right)$, where p_{inv} retains its dependence on $N(R,t)$ as in Definition 2.5. Under exponential population growth, this integral is dominated by the later portion of the interval (where population, and hence invention probability, is highest), meaning that the expected timing of an invention event under population growth is later, relative to a naive constant-population estimate based on the interval's average population, than the constant-population model would suggest, with the bias increasing with the growth rate $g(R)$ and the length of the interval.

Theorem 4.7: Population Growth Narrows, but Does Not Close, the Ceramics Case's Normalised Window

Applying the time-varying invention probability model of Definition 4.14 to the ceramics case, using palaeodemographic growth-rate estimates for the relevant Late Pleistocene-to-early-Holocene populations ($g \approx 0.02$–0.05% per year, drawn from the broader palaeodemographic literature on the terminal Pleistocene population expansion), shifts the expected invention timing later within each candidate region's feasible window relative to the constant-population estimate used in the primary analysis, which has the effect of narrowing the estimated gap between the latest-population-adjusted expected invention times across the four ceramics candidate regions. This narrowing is real but modest in magnitude — on the order of several centuries to at most one to two millennia across the full range of plausible growth-rate estimates — and does not come close to closing the multi-millennial normalised convergence window $W(\text{ceramic}) \approx 15$–$30$ established in Theorem 3.10, meaning the population-growth refinement, while a genuine improvement in modelling precision, does not alter the qualitative conclusion that

ceramic technology's polycentric emergence is robust across the full span of candidate regions.

Proposition 4.8: The Population-Growth Correction Is More Consequential for the Wheel Than for Ceramics

Because the wheel case's relevant analysis window is short (a few centuries) relative to typical Late Bronze Age population growth rates, and because the wheel case's normalised convergence window is already close to the diffusion-consistent boundary ($W(\text{wheel}) \approx 0.5$–$3$, per Theorem 3.10), the population-growth correction of Definition 4.14, while smaller in absolute chronological magnitude than for ceramics, is proportionally more consequential for the wheel case's classification: a population-growth-adjusted shift of even one to two centuries in the expected invention timing for the Pontic-Caspian source region could shift the case's normalised window non-trivially within its already-narrow and boundary-straddling range. This proportional sensitivity reinforces, from yet another angle, the wheel case's status throughout this volume as uniquely sensitive to refinements that leave the other three core cases essentially unaffected, and underscores the case for reporting the wheel's classification with the explicit uncertainty bounds this volume has consistently applied to it rather than as a settled determination.

Note 4.7: Data Limitations Preclude a Fully Quantified Population-Growth Correction for the Wheel Case

A fully quantified application of Theorem 4.7's approach to the wheel case, analogous to the ceramics calculation, would require region-specific population growth-rate estimates for the Pontic-Caspian steppe and the relevant comparison regions during the fourth millennium BCE with a precision that the current palaeodemographic literature for this period and region does not yet provide; existing estimates carry uncertainty bands wide enough that a quantified correction would add more apparent precision than the underlying data supports. The qualitative sensitivity argument of Proposition 4.8 is offered in place of a quantified correction for this reason, and improved Bronze Age steppe palaeodemographic estimates — plausibly attainable through the kind of expanded ancient-DNA sampling programme identified as a priority elsewhere in this volume (Proposition 6.6) — are identified as a precondition for closing this specific gap in the wheel case's evidentiary treatment.

Chapter 5

Methodological Apparatus of the Network Horizon Framework

5.1 Purpose of a Unified Methodological Critique

Each of the four preceding volumes deployed a specific combination of methods appropriate to its empirical subject. Volume I (horse domestication) relied primarily on ancient DNA, zooarchaeological morphometry, and lipid residue analysis of ceramic vessels associated with early equid management. Volume II (the wheel) drew on AMS radiocarbon chronology, dendrochronological cross-dating of wooden assemblages, and comparative linguistic reconstruction of wheel-and-vehicle vocabulary. Volume III (copper metallurgy) deployed lead isotope and trace element provenance analysis, metallographic examination of production sequences, and Bayesian sequence modelling of smelting-horizon dates. Volume IV (ceramics) integrated all of the above with INAA fabric analysis, computational phylogenetic glottochronology, and structural-semantic analysis of proto-language lexical domains.

The methods were introduced and defended in their case-specific contexts. What has not been provided, across the series, is a unified critical assessment: which methods carry the most discriminating power for the Diffusion Boundary Problem, which have resolution limits that constrain their utility, how they interact when deployed in combination, and what each method requires from the others in order to deliver interpretable results. This chapter provides that assessment. It is not a repetition of the methodological introductions of Volumes I–IV but an evaluation of those methods from the perspective of the unified theoretical framework developed in Chapters 1–4.

5.2 Bayesian Sequence Modelling of Radiocarbon Dates

Definition 5.1: Bayesian Radiocarbon Sequence Modelling

A *Bayesian radiocarbon sequence model* integrates n radiocarbon measurements $\{{}^{14}C_1, \ldots, {}^{14}C_n\}$ from stratigraphically ordered contexts with the IntCal calibration curve $\mathcal{C}(t)$ and a set of ordering constraints $\mathcal{O}$ derived from stratigraphic superposition, to produce posterior distributions $p(\theta_i \mid {}^{14}C_i, \mathcal{C}, \mathcal{O})$ over the calendar age θ_i of each dated event. The key advantage over conventional calibration is that the ordering constraints propagate information across the date sequence: a well-constrained date in the middle of a sequence tightens the posterior intervals of less well-constrained dates above and below it.

Discriminating power for the Diffusion Boundary Problem: high, when applied to the

chronological feasibility test of Definition 4.2. Bayesian sequence modelling can reduce 1σ calendar age uncertainty on individual events from $\pm$200–300 years (conventional calibration) to $\pm$50–100 years (sequence-modelled), which is sufficient to resolve chronological precedence between candidate invention regions separated by intervals of > 500 years. For cases like the ceramics case (inter-regional gaps of thousands of years), sequence modelling is not the limiting factor; for the wheel case (inter-regional gaps potentially as small as 200–400 years at the Pontic-Caspian/Mesopotamia boundary), it is the primary methodological bottleneck.

Resolution limits: the IntCal calibration curve has plateau regions (most notably the Hallstatt plateau at $c.$ 2,500–2,300 cal BP and a broad plateau at $c.$ 3,700–3,400 cal BP) where multiple calendar ages map to the same radiocarbon activity, producing multimodal posterior distributions that resist sequence-model compression. The earliest wheel attestations fall within or near the 3,700–3,400 cal BP plateau, which is a fundamental chronometric constraint that no methodological improvement can fully overcome.

Interaction with other methods: Bayesian sequence modelling benefits from integration with dendrochronological cross-dating (which provides precise calendar-year anchors for sequences that include timber), archaeomagnetic dating (which provides independent age constraints for fired contexts), and OSL/TL dating (which provides independent age constraints for the sedimentary matrices enclosing ceramic and metallurgical horizons). When these independent methods converge on the same sequence model, the resulting chronology has higher evidential weight than any single method alone.

Proposition 5.1: The Chronometric Resolution Threshold for Polycentric Discrimination

Polycentric emergence between two candidate regions R_i and R_j can be chronometrically distinguished from diffusion if and only if the minimum diffusion transit time $\tau_{\text{diff}}^{\min}(R_i, R_j)$ exceeds the combined chronometric uncertainty $\delta\tau_i + \delta\tau_j$ of the first-attestation dates in both regions:

$$\tau_{\text{diff}}^{\min}(R_i, R_j) > \delta\tau_i + \delta\tau_j.$$

When this condition fails — when the chronometric uncertainty envelope is wider than the minimum diffusion transit time — the chronological evidence is intrinsically unable to discriminate between models (D) and (P), regardless of the precision of the individual radiocarbon measurements. This condition defines the *chronometric resolution threshold*: cases below it require auxiliary evidence from the other three diagnostic classes (material provenance, genetics, linguistics) to resolve the Diffusion Boundary Problem.

5.3 Instrumental Neutron Activation Analysis (INAA) and Material Provenance

Definition 5.2: INAA Provenance Resolution

Instrumental Neutron Activation Analysis of ceramic fabrics measures the concentrations of ~ 30 trace and minor elements in a powdered clay sample, producing a geochemical fingerprint $\mathbf{s}(a) \in \mathbb{R}^{30}$ for each artefact a. Compositional groups are identified by multivariate clustering (Mahalanobis distance, principal component analysis, or discriminant function analysis) applied to the element concentration vectors. The provenance resolution of INAA is the minimum inter-source geographic distance $d_{\min}$ at which two source deposits produce distinguishable compositional groups, typically 50–150 km in geologically diverse regions and 200–500 km in geologically homogeneous regions.

Discriminating power: very high for the zero-contact criterion of Definition 4.3 when applied to ceramic cases, because the geographic distances between the candidate invention regions in the ceramics case (South China, Northeast Asia, Sahara, Amazonia, Near East) are all $> 5{,}000$ km — four orders of magnitude larger than INAA's resolution limit. In this regime, INAA delivers near-certain exclusion of material exchange: the probability of a false negative (genuine material exchange that INAA fails to detect) is negligible. The limiting case is the detection of contact-zone mixing between adjacent regional traditions, where artefacts produced from mixed or traded clay sources could show intermediate signatures. For the ceramics case, no inter-regional mixing signatures were detected in the Volume IV INAA programme, providing strong negative evidence for material exchange.

Resolution limits: INAA's compositional fingerprint measures the clay matrix, not the technology itself. A technology could in principle be transmitted by human movement or written instruction without any material transfer, leaving no INAA signature of contact even in a genuine diffusion scenario. The INAA zero-contact result therefore excludes *material* exchange but not *knowledge* exchange. It must be combined with the linguistic test (Definition 4.5) to assess whether knowledge transfer occurred without material transfer: genuine knowledge transfer with vocabulary borrowing would leave a loanword signature even in the absence of material exchange.

Corollary 5.1: INAA and Linguistic Diagnostics Are Complementary, Not Redundant

The INAA zero-contact result (no cross-assignment between regions) and the lexical independence result ($\phi \approx 0$, no loanword signatures in ceramic vocabulary) test two different contact channels: material and conceptual. Their joint null result — no material exchange *and* no vocabulary borrowing — provides stronger evidence for polycentric emergence than either alone, because it excludes both channels by which technology transfer could have occurred without leaving a material-composition trace. A scenario in which INAA shows no material exchange but the linguistic test shows vocabulary borrowing would be interpretively complex (knowledge-without-material transfer followed by rapid lexical nativisation) and would require independent evaluation; the joint null result obtained in the ceramics case avoids this complication.

5.4 Ancient DNA and Population Genetic Evidence

Discriminating power: ancient DNA is the most direct of the four diagnostic methods for estimating $\bar{\lambda}$, because it measures the genetic consequence of demographic contact directly rather than proxying contact through its material or linguistic traces. The genetic contact index $\gamma(R_i, R_j, t)$ of Definition 4.4 is recoverable with modern whole-genome shotgun sequencing at $0.1\times$–$1\times$ coverage, provided that adequate ancient DNA preservation exists. For the ceramics case, adequate preservation is available for South China, Northeast Asia, and the Near East; the Saharan and Amazonian cases are limited by tropical/subtropical degradation conditions that reduce DNA survival to the oldest and best-preserved specimens.

Resolution limits: ancient DNA measures *demographic* contact (gene flow between populations) rather than *technological* contact (transmission of artefacts or knowledge). The two are correlated but not equivalent: long-distance exchange networks routinely transmit objects and information without detectable gene flow between the source and recipient populations (e.g., obsidian exchange in the prehistoric Mediterranean, turquoise exchange in the American Southwest). A low genetic contact index ($\gamma \approx 0$) therefore excludes demographic contact but not all forms of technological contact; it is a strong but not conclusive negative result for the diffusion hypothesis. Conversely, a high γ confirms demographic contact but does not by

itself confirm technological transfer, since populations in contact do not necessarily share all their technologies.

Proposition 5.2: The Demographic Contact Lower Bound on Technological Contact

For any two regions R_i and R_j, the probability of sustained technological contact (transmission of a specific technology across multiple generations) is bounded above by the probability of demographic contact (sustained inter-regional movement of individuals):

$$\Pr(\text{tech contact}(R_i, R_j, t)) \leq \Pr(\text{demographic contact}(R_i, R_j, t)).$$

Therefore $\gamma(R_i, R_j, t) \approx 0$ implies $\Pr(\text{tech contact}) \leq \epsilon$ for small $\epsilon > 0$. Positive evidence for the polycentric model from the genetic diagnostic is therefore *sufficient* (low γ makes diffusion improbable) but not *necessary* (high γ is compatible with either diffusion or parallel invention in a contact zone). The genetic diagnostic has higher evidential weight as a negative test than as a positive one.

5.5 Computational Phylogenetic Glottochronology

BEAST2 Bayesian phylogenetic analysis of pottery cognate sets, developed in Chapter 9 of Volume IV, constitutes a fourth independent evidence stream for the Diffusion Boundary Problem. Its relationship to the other three methods is structurally different from the pairwise complementarities discussed in the previous sections: unlike INAA, genetics, and radiocarbon dating, glottochronology does not directly estimate $\bar{\lambda}$ or τ_{diff}. What it estimates is the age of the pottery cognate set within a given language family — which, under the Stochastic Dollo single-gain model, is an estimate of the time since the proto-language community first developed or borrowed the technology. This estimate is informative for the Diffusion Boundary Problem in two ways.

Definition 5.3: The Glottochronological Constraint on Technology Age

Let $\hat{\tau}_{\text{ling}}(\mathcal{T}, F)$ denote the BEAST2 posterior median for the age of the technology-$\mathcal{T}$ cognate set in language family F. Under the Stochastic Dollo model, $\hat{\tau}_{\text{ling}}$ is a lower bound on the time since first invention of $\mathcal{T}$ within the geographic range of F:

$$\tau_{\text{inv}}(\mathcal{T}, F) \geq \hat{\tau}_{\text{ling}}(\mathcal{T}, F).$$

The gap between $\hat{\tau}_{\text{ling}}$ and the archaeological first-attestation date $\tau_{\text{arch}}(\mathcal{T}, F)$ is the *lexical replacement gap*:

$$\Delta_{\text{lex}}(\mathcal{T}, F) = \tau_{\text{arch}}(\mathcal{T}, F) - \hat{\tau}_{\text{ling}}(\mathcal{T}, F) \geq 0.$$

A positive lexical replacement gap is consistent with both models (D) and (P) and reflects the expected rate of lexical replacement over the relevant time span. A *negative* gap ($\hat{\tau}_{\text{ling}} > \tau_{\text{arch}}$) would indicate that the linguistic estimate predates the archaeological horizon, which would be anomalous under any standard model and would require investigation of sampling bias or calibration errors.

Discriminating power: moderate but structurally distinct from the other three methods. The glottochronological method's primary contribution is lexical independence scoring (Λ_{lex} of Volume IV Definition 9.8), which quantifies the probability that the ceramic vocabulary of two language families shares a common source versus independent development. Across all five families examined in Volume IV, $\Lambda_{\text{lex}} \geq 0.85$, providing strong evidence for independent lexical development. The BEAST2 root-age estimates contribute a consistency check: if the

linguistic age of the pottery cognate set in one family post-dated the archaeological horizon in another family, this would suggest the possibility of vocabulary borrowing with technology; the absence of this pattern across all five families is consistent with polycentric emergence.

Resolution limits: the calibration horizon of $\sim 6{,}000$–$8{,}000$ years limits glottochronological utility to languages with well-dated internal calibration points in that range and to technologies whose lexical record is traceable within the horizon. For the earliest ceramic attestations ($> 10{,}000$ BP), the linguistic evidence is necessarily indirect (it constrains the age of the vocabulary, not the age of the technology itself). This limitation is not a defect of the method but a feature of the underlying linguistics: vocabulary replacement erases the oldest etymological signal, and no improvement in glottochronological methodology can recover information that the replacement process has destroyed.

5.6 The Architecture of Convergent Evidence

The four methods are not alternatives between which the researcher chooses; they are complementary components of an integrated evidential architecture. Their combination under the Bayesian framework of Definition 4.6 is optimal in the statistical sense of maximising the posterior precision of $\Pr(\bar{\lambda} < \rho^*)$ given the available data. But the practical value of the combination exceeds its formal statistical properties: each method's limitations are, in important cases, covered by another method's strengths.

Finding 5.1: Mutual Coverage of Method Limitations

Method	Primary limitation	Covered by
Bayesian ^{14}C	Calibration curve plateaux; chronometric resolution threshold	aDNA (independent age constraint on contact); INAA (independent source constraint)
INAA provenance	Material exchange only; knowledge transfer undetected	Linguistic (vocabulary borrowing test); aDNA (demographic contact)
Ancient DNA	Demographic contact only; technology transfer not guaranteed	INAA (material exchange); linguistic (vocabulary with technology)
Glottochronology	Calibration horizon; lexical replacement destroys deep signal	^{14}C (direct age); INAA (material independence)

The mutual coverage structure means that no single method failure propagates to a failure of the overall evidential case. When one method is uninformative (e.g., glottochronology for pre-$8{,}000$ BP ceramics), the other three still constrain the posterior. When two methods are equivocal (e.g., the chronometric and genetic evidence for the wheel), the overall posterior remains informative if the remaining two methods are discriminating.

Note 5.1: The Requirement for Pre-Registration of Method Combinations

The mutual coverage architecture of Finding 5.1 creates a risk of post-hoc method selection: a researcher who obtains a null result from one method could selectively report the methods that yielded positive results for polycentric emergence, assembling an apparently convergent evidential case by discarding unfavourable evidence streams. The

Network Horizon series has attempted to guard against this risk by stating, at the outset of each volume, the full suite of methods to be applied and by reporting all results including those that were equivocal or mildly unfavourable to the polycentric hypothesis (most notably the wheel's genetic evidence in Volume II). Future research applying the Network Horizon framework should pre-register the complete method suite before data collection to prevent post-hoc method selection from inflating the apparent evidential support for either model.

5.7 What the Methodological Apparatus Cannot Do

An honest methodological assessment requires explicit acknowledgement of the questions that the combined apparatus, even at its best, cannot answer.

Note 5.2: Three Questions Beyond the Method Reach

(i) Cognitive mechanism: the evidential apparatus establishes *that* independent invention occurred but says nothing about *how* cognitively similar the invention processes were across regions. Whether independent ceramic inventors in South China and the Sahara followed similar sequences of trial, error, and discovery, or whether they arrived at functionally equivalent solutions through structurally different cognitive pathways, is a question that the archaeological and linguistic evidence cannot address. Experimental cognitive archaeology and cross-cultural psychological studies of technical problem-solving offer partial approaches, but these are outside the scope of the Network Horizon framework as developed here.

(ii) Counter-factual probability: the framework establishes that polycentric emergence occurred in the four cases examined and predicts that it would occur again under similar structural conditions. But it cannot estimate how often comparable structural conditions arise without producing polycentric emergence — the counter-factual cases where all axioms are satisfied but independent invention nonetheless failed to occur within the relevant chronological window. Without this denominator, the framework cannot compute the base rate of polycentric emergence given the axiomatic conditions, which limits its predictive precision.

(iii) Sub-threshold contact: the framework is designed to detect the presence or absence of contact at the level of full technology transfer. It is less sensitive to sub-threshold contact — interactions that influenced one regional tradition in a minor way without constituting full technology transfer. Whether the South Chinese ceramic tradition had any effect on the Northeast Asian tradition via sporadic inter-regional contact that fell below the material, genetic, and linguistic detection thresholds remains genuinely uncertain. The polycentric model does not require zero contact between all pairs of regions; it requires only that full technology transfer did not occur before independent development. Minor contact below the threshold of detectability is compatible with polycentric emergence and does not invalidate the framework's conclusions.

5.8 The Problem of Method Interdependence in Multi-Source Analysis

The mutual coverage architecture of Finding 5.1 presents the four diagnostic methods as complementary and, in the ideally designed study, as independent evidence streams that can be formally combined under the Bayesian framework of Definition 4.6. In practice, however, the

four methods are not fully independent in the sense required by the statistical framework, because they share common inputs, common reference databases, and common interpretive assumptions. Identifying and managing these interdependencies is as important as assembling the evidence itself, because unidentified interdependencies inflate the apparent weight of the combined evidence beyond what the genuinely independent information content of the streams warrants.

Definition 5.4: Shared Reference Database Dependency

Two evidence streams $\mathbf{E}_a$ and $\mathbf{E}_b$ share a *reference database dependency* when both streams rely, in part, on the same reference dataset $\mathcal{D}_{\text{ref}}$ for calibration or comparison, such that errors or biases in $\mathcal{D}_{\text{ref}}$ propagate to both streams simultaneously and produce correlated errors in the resulting likelihood ratios. Examples relevant to the Network Horizon framework include: the IntCal calibration curve $\mathcal{C}(t)$, used by both the chronological stream (directly) and the glottochronological stream (indirectly, via calibrated node ages in the BEAST2 phylogeny); the INAA reference corpus for clay source discrimination, used by both the material provenance stream and (when ceramic-associated materials are used as chronometric context) the radiocarbon stream; and the ancient DNA reference panel for global population structure, used by both the genetic stream (directly) and the linguistic stream (indirectly, when genetic population structure is used to constrain the geographic interpretation of glottochronological node assignments). Reference database dependencies must be identified explicitly before the independence assumption of Definition 4.6 is applied, and correlated errors must be corrected for by either using alternative reference datasets or explicitly modelling the shared dependency in the joint likelihood function.

Proposition 5.3: The Effective Number of Independent Evidence Streams

Given a set of reference database dependencies, the effective number of independent evidence streams n_{eff} is always less than the nominal number of streams $n = 4$. For the Network Horizon series, the principal shared dependency is the IntCal calibration curve's role in both the radiocarbon chronology and the glottochronological node calibration, which induces a positive correlation between the chronological and linguistic likelihood ratios. Modelling this correlation explicitly as $\rho_{14} = 0.15$ (estimated from the overlap in chronometric anchors used by the two streams), the effective number of independent streams is approximately:

$$n_{\text{eff}} \approx 4 - 2\rho_{14} = 4 - 0.30 = 3.70.$$

The inflation factor on the nominal combined likelihood ratio is $(1+\rho_{14}) = 1.15$, meaning that Λ_{nominal} overstates the true evidential weight by approximately 15% due to this shared dependency. For the ceramics case, where $\Lambda_{\text{nominal}} > 10{,}000$, a 15% inflation is negligible; for the wheel, where $\Lambda_{\text{nominal}} = 7.1$, the corrected $\Lambda_{\text{corrected}} \approx 6.2$ — a modest change that does not alter the qualitative classification but that illustrates the importance of the correction for borderline cases.

5.9 The Role of Negative Evidence

The four diagnostic methods can each yield null results — outcomes that are consistent with both models and therefore non-discriminating. Null results are not failures; they carry evidential content in the form of absence evidence, and their proper interpretation requires explicit reasoning about what the null result implies for the posterior odds. The risk of mis-

interpreting null results is particularly acute in the Network Horizon framework, where the polycentric model's predictions are generally formulated as the absence of signals (no material exchange, no gene flow, no vocabulary borrowing) rather than the presence of positive discriminating signals.

Theorem 5.1: The Evidential Content of Null Results

A null result for evidence stream $\mathbf{E}_k$ (an observation consistent with both models) contributes a likelihood ratio of $L_k^{\text{null}} = \Pr(\text{null} \mid \bar{\lambda} < \rho^*) / \Pr(\text{null} \mid \bar{\lambda} > \rho^*)$. When the null result reflects a detection limit — the method's inability to detect small signals rather than the genuine absence of signal — the two conditional probabilities approach equality ($L_k^{\text{null}} \to 1$) and the null result is genuinely non-discriminating. When the null result reflects a true absence — the genuine non-existence of signal at the level that any method of this type could detect — the two conditional probabilities diverge in a manner that depends on the sensitivity of the method and the expected signal strength under each model. Specifically, for the material provenance zero-contact criterion (Definition 4.3), a null result at INAA resolution yields:

$$L_{\text{prov}}^{\text{null}} = \frac{\Pr(\chi \approx 0 \mid \bar{\lambda} < \rho^*)}{\Pr(\chi \approx 0 \mid \bar{\lambda} > \rho^*)} \approx \frac{1 - P_{\text{false_neg}}}{P_{\text{miss}}}$$

where $P_{\text{false_neg}}$ is the probability of obtaining $\chi \approx 0$ despite genuine material exchange (a function of the method's sensitivity) and P_{miss} is the probability of obtaining $\chi \approx 0$ under genuine non-exchange. For INAA at inter-regional distances $> 5{,}000$ km, $P_{\text{false_neg}} < 0.02$ and $P_{\text{miss}} \approx 0.95$, giving $L_{\text{prov}}^{\text{null}} \approx 47$ — a moderately strong positive likelihood ratio for polycentric emergence even from a "null" result.

This result has a practical implication: null results from high-sensitivity methods applied at large inter-regional distances are substantially more informative than null results from the same methods applied at short distances or with low sensitivity. The INAA zero-contact results of Volume IV Chapter 7, obtained at inter-regional distances of $5{,}000$–$15{,}000$ km, are therefore highly informative null results that contribute substantial likelihood ratio to the combined posterior, even though they represent the absence rather than the presence of a detected signal.

5.10 Detection of Subtle Contact: The Sub-Threshold Problem

Note 5.2 acknowledged that the multi-source evidential framework is less sensitive to sub-threshold contact — interactions that fall below the detection limits of all four methods — than to above-threshold contact. This limitation is not merely methodological but fundamental: the physical processes that leave detectable traces in ceramic compositions, skeletal DNA, radiocarbon sequences, and linguistic etymologies all have minimum intensity thresholds below which they leave no recoverable signal. Any polycentric emergence case that is in fact supported by the four methods necessarily involves the absence of detectable signals, and the question "could sub-threshold contact have occurred?" cannot, in principle, be answered with certainty by any combination of methods sensitive only to above-threshold signals.

Definition 5.5: The Sub-Threshold Contact Residual

For technology $\mathcal{T}$ and region pair (R_i, R_j) where all four evidence streams yield null results, the *sub-threshold contact residual* $\xi(\mathcal{T}, R_i, R_j)$ is defined as the maximum probabil-

ity of undetectable contact consistent with the null results:

$$\xi(\mathcal{T}, R_i, R_j) = \prod_{k=1}^{4} P_{\text{false_neg},k}(\mathcal{T}, R_i, R_j),$$

where $P_{\text{false_neg},k}$ is the probability that evidence stream k returns a null result despite genuine above-threshold contact. ξ bounds the residual uncertainty introduced by the methods' detection limits: even if all four streams return null results, there remains a probability ξ that genuine contact occurred and was missed by all streams simultaneously. For the ceramics South China–Amazonia pair, where $P_{\text{false_neg},k} \leq 0.02$ for all four streams at the relevant inter-regional distance, $\xi \leq (0.02)^4 = 1.6 \times 10^{-7}$, which is negligibly small. For the wheel's Pontic-Caspian–Mesopotamia pair, where one stream (genetics) has $P_{\text{false_neg}} \approx 0.30$ at the relevant sensitivity level, ξ is substantially larger and the sub-threshold contact residual is non-negligible.

5.11 The Chronometric Resolution Problem for Future Cases

Proposition 5.1 identified the chronometric resolution threshold — the condition under which radiocarbon-based chronology can discriminate between models (D) and (P) — and noted that it depends on the minimum diffusion transit time exceeding the combined chronometric uncertainty of the first-attestation dates. For future technologies subject to Network Horizon analysis, this threshold will vary substantially depending on the radiocarbon calibration curve's behaviour in the relevant period and on the stratigraphic quality of the earliest-attestation contexts.

Case Study 5.1: The Chronometric Situation for Textile Production

Textile production (Finding 6.1's second candidate technology) presents a chronometric situation intermediate between ceramics and the wheel. The earliest secure attestations cluster in three date ranges: Near Eastern linen contexts at $c.$ 9,000–8,500 BP (within the early Holocene atmospheric ^{14}C plateau), Chinese silk contexts at $c.$ 7,000–6,500 BP (outside the plateau, with good calibration precision), and South American cotton contexts at $c.$ 6,000–5,500 BP (in a period of good calibration precision). The minimum diffusion transit time between the Near East and China is estimated at $c.$ 4,000 years at $v_{\text{max}} = 200$ km/century over the 8,000 km route through Central Asia. The observed inter-attestation spread of $\approx$ 2,000 years between the Near Eastern and Chinese attestations is therefore shorter than $\tau_{\text{diff}}^{\text{min}}$, placing the Near East–China pair above the chronometric resolution threshold of Proposition 5.1. The chronological infeasibility criterion does not apply to the Near East–South America pair, where the spread of $\approx$ 3,000 years exceeds the minimum transit time over the 20,000 km trans-oceanic route, making independent invention the only chronologically feasible explanation. The textile case therefore has a mixed chronometric situation: clear chronological infeasibility for some pairs, ambiguity for others, and a requirement for the full four-stream analysis to resolve the ambiguous cases.

5.12 The *Chaîne Opératoire* as a Falsifying Tool

The most fine-grained structural analysis available for the independent-invention hypothesis is the *chaîne opératoire* approach introduced in Case Study 2.3 and applied extensively in

Volume IV. Its diagnostic power for the Diffusion Boundary Problem derives from the fact that transmitted technical traditions preserve non-functional choices — specific sequences of actions, specific material selections, specific morphological conventions — that are unnecessary for the technology's function and would not be independently reproduced by an inventor starting from scratch. The presence of shared non-functional choices across regional traditions is therefore positive evidence for transmission; their absence is positive evidence for independence.

Definition 5.6: The Non-Functional Choice Divergence Index

For two regional ceramic (or metallurgical, equid-management, etc.) traditions $\mathcal{T}_i$ and $\mathcal{T}_j$, define the *non-functional choice set* $\mathcal{NF}(\mathcal{T}_i)$ as the set of technical choices made in the production of $\mathcal{T}_i$ that are not required by any functional criterion in the technology's specification. The *non-functional choice divergence index* is:

$$D_{\mathcal{NF}}(\mathcal{T}_i, \mathcal{T}_j) = 1 - \frac{|\mathcal{NF}(\mathcal{T}_i) \cap \mathcal{NF}(\mathcal{T}_j)|}{|\mathcal{NF}(\mathcal{T}_i) \cup \mathcal{NF}(\mathcal{T}_j)|},$$

the Jaccard dissimilarity between the non-functional choice sets of the two traditions. Under transmission from a common template, $D_{\mathcal{NF}} < D_{\mathcal{NF}}^{\text{drift}}$ (constrained below the drift rate from the template); under independent invention, $D_{\mathcal{NF}} \approx 1 - P_{\text{coincidence}}$ where $P_{\text{coincidence}}$ is the probability of independently arriving at the same non-functional choice, which is low for most choices (different clay tempering strategies, different coil-build versus paddle-and-anvil forming techniques, different firing atmosphere management approaches).

Finding 5.2: Non-Functional Choice Divergence for the Ceramics Case

Applying Definition 5.6 to the ceramic traditions of the five independently attested ceramic regions (Volume IV Chapter 6), the pairwise non-functional choice divergence indices $D_{\mathcal{NF}}$ are reported below. *Values are derived from petrographic, morphological, and forming-technique comparisons assessed by blind coding of regional assemblages (Volume IV Chapter 6); the index ranges from 0 (identical production choices) to 1 (completely divergent). The table reports the qualitative divergence band and whether the pattern is consistent with a contact or an independent-origin model for each pair.*

Region pair	$D_{\mathcal{NF}}$ band	Interpretation
South China – NE Asia/Japan	high (> 0.80)	Independent; modest geographic proximity
South China – Near East	very high (> 0.88)	Strongly independent; distinct functional drivers
South China – Sahara	very high (> 0.95)	Strongly independent; no evidential basis for conta
South China – Amazonia	very high (> 0.97)	Strongest divergence pair
NE Asia/Japan – Near East	very high (> 0.87)	Strongly independent
NE Asia/Japan – Sahara	very high (> 0.93)	Strongly independent
NE Asia/Japan – Amazonia	very high (> 0.96)	Strongly independent
Near East – Sahara	very high (> 0.86)	Strongly independent; distance and environment
Near East – Amazonia	very high (> 0.94)	Strongly independent
Sahara – Amazonia	very high (> 0.95)	Strongly independent

All pairwise indices reach or exceed the "high" divergence band; the lowest value (South China – NE Asia/Japan) corresponds to the geographically closest pair and the pair that also shows the smallest genetic contact index (Finding 4.1), suggesting that some degree

of sub-threshold contact or shared ecological context may account for the modestly higher non-functional choice overlap between these two traditions. The pattern of distance-decay in $D_{\mathcal{NF}}$ is consistent with a polycentric model in which independent invention is accompanied by local elaboration that diverges more rapidly with geographic distance from any potential common influence.

5.13 Integrating Ethnoarchaeological Evidence

The four primary diagnostic methods of Chapter 4 are all applied to the prehistoric archaeological record itself. A complementary line of evidence that has been used selectively but not systematically in the preceding volumes is ethnoarchaeological: direct observation of traditional craft production in contemporary communities where the technology has been practised continuously, providing a living analogue for the production sequences, knowledge transmission patterns, and community learning dynamics that underlie the prehistoric record.

Proposition 5.4: The Scope and Limits of Ethnoarchaeological Analogy

Ethnoarchaeological evidence from contemporary ceramic-producing communities (Gosselain 1992; Roux 2019; Livingstone Smith 2001) provides calibration data for three parameters of the Network Horizon model: (i) the minimum viable threshold $\theta(\mathcal{T})$, derived from experimental studies of naive learners' acquisition times for basic ceramic forming skills; (ii) the transmission coefficient $\alpha(\mathcal{T})$, estimated from the fraction of contact events between a ceramic-producing and a non-ceramic-producing community that result in technology transfer within one generation; and (iii) the non-functional choice drift rate, which governs the rate at which independent ceramic traditions diverge in their non-functional technical choices over generations of transmission. Ethnoarchaeological analogy is limited by the structural differences between contemporary ceramic-producing communities and their prehistoric analogues: contemporary communities operate in a world of much denser contact networks, much more abundant alternative container technologies, and much higher population densities than the late Pleistocene and early Holocene communities in which the ceramics polycentric emergence is argued to have occurred. The ethnoarchaeological calibration must therefore be adjusted downward (for lower contact densities) and upward (for higher functional pressure from the absence of alternatives) before application to the prehistoric cases.

5.14 Inter-Rater Reliability in Axiom Scoring

The composite axiomatic scores $\Sigma(\mathcal{T})$ reported throughout the series depend on expert judgement at several stages: assessing whether a given body of archaeometric, genetic, or chronological evidence satisfies a given axiom's evidential requirements, and to what degree. This dependence on judgement raises a methodological question not yet directly addressed: how reliable are these scores across independent assessors, and to what extent does the framework's conclusions depend on assessor-specific interpretive choices rather than on the evidence itself.

Definition 5.7: Inter-Rater Reliability for Axiom-Satisfaction Scoring

For a set of n technology–region cases scored independently by m assessors using the axiom-satisfaction protocol of Definition 2.4, inter-rater reliability is quantified using the intraclass correlation coefficient $\mathrm{ICC}(2,1)$, appropriate for a fully crossed design in which all assessors score all cases and both assessor and case are treated as random effects. An ICC above 0.75 is conventionally taken to indicate good reliability; values between 0.5 and 0.75 indicate moderate reliability requiring caution in interpreting close classification calls; values below 0.5 indicate that the scoring protocol requires further operationalisation before its outputs can be treated as more than provisional.

Worked Example 5.5: What an Axiom II Reliability Check Would Look Like

No inter-rater reliability study has in fact been carried out for this series: the axiom scores reported in Chapters 3 and 6 are the lead investigator's own point estimates, not the output of multiple independent assessors. The following is a hypothetical illustration, constructed to show what such a study would need to look like and what an ICC-based reliability report would mean if one were conducted, not a result obtained from real data. Suppose three independent assessors with backgrounds in experimental archaeology scored Axiom II (minimum viable threshold) for the four core cases plus the iron-smelting negative control of Worked Example 1.3. A design of this kind, analysed with the intraclass correlation coefficient $\mathrm{ICC}(2,1)$ of Definition 5.7, could in principle return a value indicating good agreement (conventionally, above 0.75) for some axioms and only moderate agreement (in the 0.5–0.75 band) for others — plausibly including Axiom IV (demographic isolation), where translating sparse ancient-DNA admixture estimates into a single contact-density score involves genuine interpretive latitude that different assessors could reasonably resolve differently. This is offered as a plausible scenario, not a finding.

Note 5.4: Reliability as an Unmet Methodological Need, Not a Completed Check

Worked Example 5.5 must not be read, by the "worked example" label or otherwise, as reporting data: it is a hypothetical scenario sketched to make the methodological gap concrete, and no assessors, no ICC computation, and no reliability data of any kind exist for this series at the time of writing. This is the single most consequential unmet methodological need identified anywhere in this volume, because every classification reported in Chapters 3 and 6 currently rests on one analyst's judgement with no inter-rater check of any kind. A genuine reliability programme would require: a panel of independent domain specialists recruited specifically for the task; a pre-registered scoring protocol applied blind to which technology and which hypothesis a given evidence packet is meant to support; scoring of all four core cases plus suitable negative controls; and computation of $\mathrm{ICC}(2,1)$ from the resulting score matrix, reported with the raw per-assessor scores available for independent audit. Until such a programme is carried out and its data published alongside this series, every composite score $\Sigma(\mathcal{T})$ in this volume should be read as a single, unreplicated expert judgement, and any apparent precision in its reported value should be discounted accordingly.

5.15 Sensitivity of Conclusions to Expert Disagreement Resolution Procedures

When independent assessors disagree on an axiom score, some procedure is needed to arrive at a single value for use in the composite score $\Sigma(\mathcal{T})$. This section examines how sensitive the series's overall conclusions are to the choice of disagreement-resolution procedure.

Proposition 5.5: Aggregation Rule Choice Is, in Principle, Unlikely to Be Outcome-Determining for the Four Core Cases — An Untested Claim

No multi-assessor data exist for this series (Note 5.4), so no aggregation rule — simple averaging, median selection, or structured deliberation toward consensus — has actually been compared on real disagreement data, and the claim of this proposition is accordingly a structural argument, not an empirical result. The argument is this: for the four core cases, the lead investigator's composite scores, with the partial exception of the wheel, lie sufficiently far from the Regime I/Regime II boundary ($\Sigma = 0.5$) that a modest spread between assessors on any individual axiom — of the order suggested by the hypothetical scenario of Worked Example 5.5 — would plausibly leave the regime classification unchanged under any of the three aggregation rules. This is a plausibility argument resting on distance-from-threshold, not a demonstrated robustness result; it does not establish that aggregation-rule choice is immaterial in general, and cases closer to the boundary (such as prospective cases evaluated under the Chapter 6 methodology) should be assumed more sensitive to this methodological choice until real multi-assessor data say otherwise.

5.16 The Role of Disconfirming Expert Opinion

A methodologically important question, distinct from inter-rater reliability in the statistical sense, concerns how the framework should treat cases where a recognised subject-matter expert offers a considered judgement that conflicts with the framework's composite-score classification, even after accounting for ordinary measurement disagreement of the kind documented above.

Corollary 5.6: Expert Disagreement as a Distinct Category from Measurement Noise

Expert disagreement that persists after accounting for measurement-level inter-rater variation (Definition 5.7) should be treated as a distinct epistemic category from ordinary scoring noise, since it may indicate either a genuine evidential consideration not captured by the existing axiom-scoring protocol (in which case it should prompt protocol revision, following the logic of Corollary 2.4's parsimony constraint on axiom proliferation) or a difference in background theoretical commitments not resolvable by appeal to the evidence alone (in which case the disagreement should be reported transparently rather than adjudicated by the framework's own internal procedures, which are not designed to resolve disputes about background theory). The Botai horse-domestication debate discussed at length elsewhere in the series exemplifies the first category: sustained expert disagreement about the bit-wear and lipid-residue evidence ultimately prompted refinement of how Axiom II's threshold scoring incorporates contested archaeozoological indicators, an instance of the framework responding constructively to expert disagreement rather than overriding it by fiat.

5.17 Negative Evidence Revisited: Formal Treatment of Absence

Section 5's discussion of negative evidence treated absence of expected diffusion-consistent traces as informative under specified conditions. This section provides the formal Bayesian treatment underlying that informal discussion, making explicit the assumptions required for absence of evidence to function as evidence of absence within the framework.

Theorem 5.6: Conditions for Informative Negative Evidence

Let $\mathbf{E}^{-}$ denote the absence of a specific evidential trace (for example, the absence of detectable INAA fabric-composition homogeneity across two regions under comparison) that would be expected with probability p_1 under the diffusion hypothesis and probability p_2 under the polycentric hypothesis. The absence $\mathbf{E}^{-}$ is informative (contributes a Bayes factor different from 1) if and only if $p_1 \neq p_2$, and is informative *in favour of polycentric emergence* specifically when $p_1 > p_2$, that is, when the trace would have been more likely to be detected under diffusion than under independent origin. This condition holds for fabric-composition homogeneity (expected under diffusion via shared raw-material sourcing or shared technical recipe; not expected under independent origin from locally idiosyncratic materials) but does not hold automatically for all forms of absence: the absence of a textual record of contact, for instance, is uninformative for the deep-prehistoric cases addressed by this series, since no textual record exists or would be expected to exist under either hypothesis, so $p_1 = p_2 = 0$ and Theorem 5.6's informativeness condition fails. Researchers applying the negative-evidence reasoning of this series to new cases must verify, for each specific absent trace under consideration, that the detection probabilities genuinely differ across the two hypotheses before treating the absence as evidentially informative.

5.18 Computational Reproducibility and the Audit Trail

The methodological apparatus developed across this chapter — the multi-source evidence-integration framework and the negative-evidence reasoning of Theorem 5.6 — depends at several stages on specific numerical inputs (population estimates, calibration curve parameters, reference database compositions) drawn from the primary literature, and, in the radiocarbon and genetic sub-fields specifically, on third-party software implementations (Bayesian sequence-modelling and phylogenetic pipelines such as those discussed in Section 5.2 and 5.4) whose reproducibility is a precondition for the framework's claims to be independently verifiable rather than merely independently assertable. The inter-rater reliability protocol of Definition 5.7 is not part of this audit-trail discussion, since, as Note 5.4 makes explicit, it has not yet been applied to real data and there is accordingly no ICC computation of this series' own to audit. This section addresses the computational reproducibility infrastructure underlying the series' quantitative claims about the primary evidence it synthesises.

Definition 5.8: The Computational Audit Trail

A *computational audit trail* for a quantitative claim Q derived within the Network Horizon framework consists of: (i) the specific dataset or datasets used as input, with version identifiers or access dates sufficient to locate the exact data state used (relevant for datasets, such as reference DNA panels and the IntCal calibration curve, that are periodically revised); (ii) the specific software, including version numbers, used to process the data; (iii) the specific parameter settings (priors, thresholds, random seeds where applicable) used in the analysis; and (iv) a record sufficient for an independent researcher

to reproduce Q from (i)–(iii) without requiring access to the original researcher's working files beyond what is documented. An audit trail is *complete* if all four components are specified with sufficient precision for independent reproduction; it is *partial* if one or more components are specified only in general terms (for instance, citing a software package without a version number, or a dataset without an access date).

The four core cases of this series vary in audit trail completeness, reflecting both the heterogeneity of the underlying primary literature (much of which predates current reproducibility norms in archaeology and genetics) and the practical constraints of synthesising results across dozens of primary sources rather than conducting all analyses from raw data within a single unified computational pipeline.

Proposition 5.6: Audit Trail Completeness Correlates with Recency of the Underlying Primary Research

Across the citations underlying the four core cases' axiom-satisfaction scores, audit trail completeness (Definition 5.8) is markedly higher for ancient-DNA-derived claims published after approximately 2018, reflecting the increasing standardisation of bioinformatic reproducibility practices (deposited raw sequencing reads, versioned analysis pipelines, publicly archived parameter files) in that sub-field over the past decade, than for claims derived from earlier archaeometric and archaeological literature, where dataset versioning and computational parameter documentation were less consistently practised as a matter of disciplinary norm. This pattern implies that the series' more recently incorporated genetic evidence (much of which bears on Axiom IV, as discussed in Note 3.2) is, on average, more independently verifiable than its longer-standing archaeometric and archaeological evidence (bearing primarily on Axioms I, II, and III), a methodological asymmetry that mirrors, from a reproducibility rather than an evidentiary-strength standpoint, the trajectory-reversal pattern already noted in Note 3.2.

Theorem 5.7: Reproducibility Limitations Do Not Undermine the Multi-Source Robustness Argument

The multi-source robustness argument of Finding 5.1 — that convergent evidence across independent streams provides stronger support than any single stream alone — does not require that every contributing stream meet the full reproducibility standard of Definition 5.8; it requires only that the streams' underlying evidential claims be independently checkable in principle, even if full computational reproduction of every intermediate analytical step has not, in every case, been documented to current standards. The distinction matters because it separates two different senses in which a finding can be "verified": re-deriving a numerical result from documented raw data and code (full computational reproducibility), versus independently re-examining the underlying physical or genetic evidence using the same or comparable methods to check whether a qualitatively similar conclusion is reached (independent replication in the broader scientific sense). The four core cases of this series have, in every instance bearing on a load-bearing axiom score, been subject to independent replication in the second sense — multiple research groups, often using different specific samples and sometimes different specific methods, reaching concordant conclusions — even where full audit-trail documentation in the first, narrower sense is incomplete for some of the older contributing studies.

Note 5.5: A Recommendation for Future Volumes and Companion Studies

Notwithstanding Theorem 5.7's argument that incomplete audit-trail documentation does not undermine the series' substantive conclusions, completing the audit trail for the framework's own derived quantities — the composite score calculations, the diffusion-velocity estimates, the Bayesian combination of likelihood ratios — is a tractable and worthwhile undertaking that does not depend on retroactively documenting decades of underlying primary literature to modern standards. The authors recommend, as a priority for the companion technical materials accompanying future volumes of this series, the publication of a versioned computational repository containing the specific parameter values, calculation scripts, and intermediate outputs underlying every numbered finding in this volume, so that the framework's own derivations — as distinct from the primary evidence they synthesise — meet the full standard of Definition 5.8. This is a modest undertaking relative to the scope of the primary research synthesis and would substantially strengthen the series' claim to falsifiability in the practical, not merely the in-principle, sense emphasised throughout Chapter 6.

Corollary 5.7: The Distinction Between Primary-Evidence Reproducibility and Framework-Derivation Reproducibility

Critics evaluating the Network Horizon framework's reproducibility should distinguish two separate questions that the discussion of this section has kept analytically distinct: whether the underlying primary archaeological, genetic, and linguistic evidence synthesised across Volumes I–IV is independently verifiable (a question whose answer varies study by study and is, per Theorem 5.7, generally satisfied in the relevant replication sense for load-bearing claims), and whether the specific quantitative derivations original to this volume — the composite scores, the diffusion-velocity calculations, the Bayesian likelihood combinations — are computationally reproducible from documented inputs (a question whose answer is, per Note 5.5, currently incomplete and identified as a priority for remediation). Conflating these two questions risks either over-crediting the framework's own reproducibility on the strength of well-replicated primary evidence, or over-discounting well-replicated primary evidence on the grounds of the framework's own incomplete audit trail; both errors are avoided by keeping the two reproducibility questions separate, as this section has attempted to do.

5.19 Comparative Evaluation of Alternative Statistical Combination Frameworks

The Bayesian and Dempster–Shafer combination approaches used elsewhere in this chapter to integrate the four evidence streams are not the only statistical frameworks available for multi-source evidence integration, and a brief comparative evaluation of the road not taken clarifies why the Bayesian approach was preferred for this series' purposes.

Definition 5.9: Alternative Multi-Source Integration Frameworks

Beyond the Bayesian likelihood-ratio combination and Dempster–Shafer belief-function approaches used in this volume, at least two further statistical frameworks are available for integrating evidence across the four diagnostic streams: *meta-analytic effect-size pooling*, which treats each stream's evidence as an independent estimate of a common underlying effect size and combines them via inverse-variance weighting, a standard approach in evidence-based medicine and the social sciences; and *frequentist multiple-*

hypothesis testing with appropriate correction for the number of streams tested, which treats each stream as a separate test of the null hypothesis of no polycentric emergence and combines p-values via methods such as Fisher's combined probability test.

Proposition 5.7: Meta-Analytic Pooling Is Poorly Suited to Qualitatively Heterogeneous Evidence Streams

Meta-analytic effect-size pooling presupposes that the different evidence streams estimate a common underlying quantity on a shared scale, an assumption well suited to combining multiple studies that each estimate, for instance, a treatment effect in comparable units. The four diagnostic streams of this volume's evidentiary framework — chronological feasibility ratios, archaeometric cross-assignment rates, genetic admixture proportions, and glottochronological lexical independence scores — are not naturally expressed on a common effect-size scale, and forcing them onto one (for instance, by standardising each to a z-score relative to its own null distribution before pooling) discards information about the qualitatively different evidential content of each stream that the Bayesian likelihood-ratio framework preserves more naturally, since likelihood ratios from qualitatively different evidence types combine multiplicatively under standard Bayesian logic without requiring a forced common scale. This consideration, rather than any deficiency specific to meta-analytic methods in their native domain, motivated the Bayesian framework's adoption as the primary integration method throughout this series.

Theorem 5.8: Frequentist Multiple-Testing Correction Would Be More Conservative Than Necessary for This Application

Fisher's combined probability test and related frequentist multiple-hypothesis-testing frameworks are designed for settings where the analyst wishes to control the family-wise error rate or false discovery rate across many simultaneous hypothesis tests, a concern most pressing when a large number of hypotheses are tested and only a small fraction are expected to be true. The present application — combining exactly four pre-specified evidence streams bearing on a single, pre-specified hypothesis for a single, pre-specified technology — does not exhibit the large-multiple-comparisons structure that motivates frequentist correction procedures, and applying such corrections here would impose an unnecessarily conservative penalty relative to the Bayesian framework's direct accumulation of evidential weight via likelihood-ratio multiplication, which already incorporates the appropriate discounting for non-independence via the correlated-evidence-stream analysis of Proposition 5.3. This is a structural argument for why the Bayesian framework was adopted as the primary integration method, not a report of an empirical comparison: Fisher's method has not in fact been run as a cross-check against the four core cases' chronological feasibility tests, and no claim is made here that it would, if run, agree with the Bayesian conclusions in every instance — that is a separate, empirical question left open for future work.

Corollary 5.8: Framework Choice Is Unlikely to Be Outcome-Determining for the Four Core Cases — A Theoretical Expectation, Not a Tested Result

The comparative evaluation of this section is a theoretical argument about why Bayesian likelihood-ratio combination was adopted, not an empirical demonstration that it, Dempster–Shafer belief functions, and frequentist combined-probability testing in fact agree when run on the four core cases: as Theorem 5.8 makes explicit, the frequentist cross-check has not actually been performed, and the Dempster–Shafer approach, while

discussed structurally in this chapter, has likewise not been independently re-run against the primary data to confirm it yields the same classifications. The expectation that the three frameworks would concur — classifying ceramics, copper metallurgy, and horse domestication as strongly polycentric and identifying the wheel as the boundary case — rests on the structural arguments of Propositions 5.7 and the discussion above, not on a completed multi-framework comparison. Carrying out that comparison on real data is identified here as a further item for the methodological agenda alongside the inter-rater reliability programme of Note 5.4, not as a robustness result already in hand.

Chapter 6

Falsifiability and Predictive Scope

6.1 A Theory Without Falsifiable Predictions Is Not a Theory

The Network Horizon model's axiomatic structure (Chapter 2) and its retrospective application to four cases (Chapter 3) together constitute a coherent and internally consistent framework. But internal consistency and retrospective fit are not sufficient to establish the model as a scientific theory rather than a post-hoc narrative. For the model to earn the status of a theory, it must generate predictions about cases not used in its construction — predictions that could, in principle, be falsified by future empirical work. This chapter derives those predictions.

Two classes of predictions are distinguished: *structural predictions* about which currently unexamined technologies should show polycentric signatures given their axiomatic properties, and *diagnostic predictions* about what specific evidential patterns should be found if polycentric emergence occurred. Both classes are testable; the structural predictions by expanding the corpus of technologies subjected to Network Horizon analysis, the diagnostic predictions by improving the resolution of the methods discussed in Chapter 5.

Note 6.0: Calibration Set Versus Test Set — A Distinction This Chapter Must Honour

Before any prediction is offered, the distinction the rest of this chapter depends on needs to be made explicit and then actually respected. The four core cases (ceramics, copper, horse domestication, the wheel) are the *calibration set*: the axioms of Chapter 2 were extracted from, and the σ-score elicitation procedure was developed against, exactly these four cases. Any case drawn from the calibration set cannot, by construction, constitute independent evidence for the model — a model will always fit the data it was built to fit. A genuine test requires a case that played no role in shaping the axioms, the scoring procedure, or the choice of which evidentiary streams to weight.

Agriculture, discussed immediately below as Finding 6.1's first case, fails this test in a way that needs to be stated rather than elided: the polycentric emergence of plant domestication across several world regions was already a well-established, independently documented finding in archaeobotany and genetics (Zeder 2011; Fuller et al. 2011; Larson et al. 2014) at the time this framework was constructed, and it is a sufficiently well-known case that it plausibly informed the intuitions behind Axioms I and III even if it was not formally included in the four-case calibration corpus. Presenting the model's retrodiction of an already-known result as confirmation, as Proposition 6.1 below does, is therefore weaker evidence than it may appear, and Proposition 6.1 is annotated accordingly. The genuinely informative tests in this chapter are the predictions for textiles and iron smelting, which involve currently under-analysed or contested evidentiary records and were not used, even informally, in constructing the axioms. The remainder of this chap-

ter should be read with this distinction in view: confirmatory restatements of already-known polycentric cases (agriculture) demonstrate consistency, not predictive validity; genuinely prospective predictions on unanalysed cases (textiles, iron smelting, and the worked cases later in this chapter) are the actual locus of falsifiability claimed for this framework.

6.2 Structural Predictions: Technologies Expected to Show Polycentric Emergence

The axiomatic framework of Chapter 2 generates technology-specific predictions about $\Sigma(\mathcal{T})$ without requiring prior empirical analysis of the technology in question. A technology's composite polycentric score can be estimated *a priori* from its material requirements (Axiom I), production complexity (Axiom II), functional pressure distribution (Axiom III), and the contact-network density at its likely emergence period (Axiom IV). The following cases are examined as predictions of the model.

Finding 6.1: Predicted Axiomatic Profiles for Three Untested Technologies

Independent plant domestication (Neolithic agriculture): The archaeological record already documents multiple independent centers of plant domestication: the Fertile Crescent ($c.$ 11,000–10,000 BP), the Yangtze and Yellow River basins ($c.$ 9,000–7,500 BP), New Guinea highlands ($c.$ 9,000 BP), the Mexican highlands ($c.$ 9,000–8,000 BP), the Andes ($c.$ 8,000 BP), and the eastern North American woodlands ($c.$ 5,000–4,000 BP). The Network Horizon model predicts this pattern: wild plant populations suitable for domestication are widely distributed (Axiom I: *strong*, comparable to ceramics); the minimum viable threshold for initial domestication is achievable by any sedentary or semi-sedentary forager group already engaged in intensive plant exploitation (Axiom II: *strong*); the functional pressure recurs wherever population density exceeds the reliable carrying capacity of wild foraging (Axiom III: *strong*); and the contact densities between all six domestication centres at their respective emergence dates are well below any plausible ρ^* for agriculture (Axiom IV: *strong*). The overall axiomatic profile is *uniformly strong* across all five axes, comparable to ceramics, consistent with the strong empirical support for agricultural polycentrism that already exists as independent scholarly consensus. As Note 6.0 and Proposition 6.1 discuss below, this consistency is a sanity check on the model's calibration, not an independent confirmation, since agriculture's polycentric status was already established knowledge prior to this framework's construction.

Textile production (weaving and spinning): Archaeological evidence for independent textile traditions exists in the Near East ($c.$ 9,000 BP for linen), China ($c.$ 7,000 BP for silk), and South America ($c.$ 6,000 BP for cotton). Plant fibres suitable for spinning are distributed across all inhabited regions (Axiom I: *strong*); spinning and simple over-under weaving require dexterity and patience but no specialised tool beyond a spindle whorl and a simple loom (Axiom II: *strong*); the functional pressure (thermal insulation, load-carrying, social signalling) recurs universally (Axiom III: *strong*); contact densities at emergence dates are low for the trans-oceanic pairs (Near East/China/South America) (Axiom IV: *strong*). The composite axiomatic profile places textile production above the intermediate range, predicting moderate-to-strong polycentric evidence for textile production that has not yet been subjected to the full multi-source evidentiary analysis applied in this series.

Iron smelting: Iron smelting requires sustained temperatures above 1,200°C with carburation control — a substantially higher minimum viable threshold than copper smelting (Axiom II: *moderate–weak*, the binding constraint, falling clearly below copper smelting). Iron ore deposits are widely distributed (Axiom I: *strong*) and functional pressure is high (Axiom III: *strong*), but the Axiom II deficiency dominates. The overall predicted axiomatic profile places iron smelting below the wheel in polycentric proneness, predicting that the combined evidential weight should favour the polycentric model only weakly or equivocally — significantly weaker than even the wheel.

Proposition 6.1: The Agriculture Case as a Consistency Check, Not an Independent Test

The polycentric emergence of agriculture is not a prediction to be tested but a finding already established by the independent research programmes of Zeder (2011), Fuller et al. (2011), Larson et al. (2014), and the many regional studies synthesised in those reviews. The Network Horizon model's retrodicted composite profile for agriculture falls comfortably in the *strong* polycentric band, consistent with this prior finding without requiring modification of the model's parameters, and this consistency is worth reporting. But it should not be reported, as an earlier formulation of this proposition did, as the model "independently predicting" or being "confirmed by" a fifth case outside its calibration set: agriculture's polycentric status was general knowledge in archaeobotany well before this framework's axioms were drafted, and the axioms were not constructed in ignorance of it. What this case actually demonstrates is the weaker but still useful claim of Note 6.0: the framework's parameter estimates, elicited from the four core cases, do not produce an obviously wrong answer when pointed at a fifth, well-understood case. This is a basic sanity check that a badly mis-specified model could fail; passing it is necessary but far from sufficient grounds for confidence in the model's predictive validity. The textile and iron-smelting cases below carry correspondingly more evidential weight precisely because they are not subject to this objection.

6.3 Diagnostic Predictions: What Future Evidence Should Find

Beyond structural predictions about which technologies should show polycentric signatures, the model generates diagnostic predictions about what specific patterns the evidence should exhibit *if* polycentric emergence occurred. These predictions are falsifiable: if future evidence finds the predicted patterns absent, the model is challenged; if it finds them present, the model is corroborated.

Finding 6.2: Twelve Falsifiable Diagnostic Predictions of the Network Horizon Model

The following predictions follow directly from the axiomatic structure developed in Chapter 2 and the evidential framework of Chapters 4–5. Each is stated in a form that specifies the evidence that would confirm or disconfirm it.

P1 For any technology with predicted $\Sigma > 0.5$, the material provenance zero-contact criterion (Definition 4.3) should be satisfied: $\chi(R_i, R_j) \approx 0$ for all pairs of candidate regions. *Falsified by*: significant cross-assignment rates between any two candidate regions.

P2 For any technology with predicted $\Sigma > 0.5$, the genetic contact index $\gamma(R_i, R_j, t)$ at the relevant emergence dates should be < 0.05 for all trans-continental pairs. *Falsified by*: aDNA evidence for substantial gene flow between candidate regions within

the relevant chronological window.

P3 For any technology with a *strong* or *moderately strong* overall axiomatic profile (predicted polycentric), the lexical independence score Λ_{lex} between the corresponding proto-language technology-vocabulary domains should be clearly above the loanword signature threshold for all pairs, and loanword signatures should be absent. *Falsified by*: cognate relationships or loanword signatures in the core technology vocabulary of two or more candidate language families.

P4 The F_1/F_2 register ratio of Definition 9.7 of Volume IV should differ systematically between polycentric-Regime-I technologies (dominated by processing-function morphemes, $F_1/F_2 > 1.5$) and diffusion-Regime-II technologies (with higher storage or display-function morphemes, F_1/F_2 variable). *Falsified by*: a polycentric technology with $F_1/F_2 < 1$ in its proto-language ceramic domain.

P5 For technologies at the Regime I/II boundary (those whose axiomatic profile places them in the *weak* overall band, comparable to the wheel), the posterior distribution of $r = \bar{\lambda}/\rho^*$ should straddle $r = 1$, with the posterior clearly uncertain rather than decisive. *Falsified by*: a boundary-regime technology for which the posterior is decisively on one side of the boundary.

P5b For any monocentric diffusion model of wheeled-vehicle technology that identifies a specific geographic locus of origin (the Pontic-Caspian steppe, the Trypillia–Yamna contact zone, or any other single region), the earliest secure attestation dates at that locus must predate attestations in all other regions by an interval exceeding the minimum diffusion transit time estimated from Bronze Age contact-network parameters (Definition 4.2). Given the IntCal20 calibration plateau in the relevant interval ($c.\ 4{,}700$–$4{,}900$ BP, with effective chronometric resolution ± 150–200 years), a monocentric model requires a chronological lead at the proposed origin of at least 300–400 years relative to the nearest attested secondary region. *Falsified by*: the current Ward–Wilson test result ($\chi^2 = 3.4$, two degrees of freedom, critical threshold 5.99), which shows that contemporaneity of the Bronocice, Yamna funerary, and Uruk IVa attestations is not rejected at $\alpha = 0.05$. The monocentric model is not falsified by the *absence* of a significant gradient; it is placed under evidential pressure by it. Future confirmation would require radiocarbon measurements with systematic biases sufficiently large to have produced the flat gradient from an actually steep one, or stratigraphic reinterpretation that relocates one or more of the three key attestations by 300+ years. The current evidence does not support either.

P5c For the specific monocentric co-invention model proposed by Haarmann (2016) — in which wheeled-vehicle technology emerges from bilingual technological collaboration in the Trypillia–Yamna contact zone with lexical contributions from a paleo-European substrate — the following joint prediction must be satisfied: (a) the PIE lexicon of wheeled transport should contain at minimum one term satisfying the three-criterion substrate-evidence protocol (pre-IE etymology documented, geographic distribution consistent with the Trypillia zone, chronologically appropriate semantic field), and (b) the earliest wheel attestations should show a geographic concentration within the proposed contact zone that precedes attestations outside it by an archaeologically discriminable interval. *Partially falsified*: condition (a) is not met by any of the eleven candidate substrate terms in Haarmann's corpus (Volume II, Section 8.2.3, score 0/11 under the pre-specified three-criterion protocol); condition (b) is not established at current chronometric resolution (P5b above).

Neither sub-prediction is independently verified; their joint probability of being simultaneously correct given current evidence is the product of two low probabilities, placing the co-invention model in a very weak evidential position without requiring its categorical falsification.

P6 Structural divergence between regional instances of a polycentric technology (Axiom V) should be positively correlated with $\tau_{\text{diff}}^{\min}$ between those instances: more isolated regions should show more divergent technical solutions. *Falsified by*: high structural similarity between the most isolated pairs of regional instances.

P7 For textile production, the Network Horizon model predicts INAA-style fabric analysis of early fibre-processing tools and weights should find no cross-assignment between Near Eastern, Chinese, and South American assemblages at their respective emergence dates. *Falsified by*: compositional signatures indicating shared material sources between any two centres.

P8 For iron smelting, the model predicts that the Bayesian posterior $\Pr(r < 1)$ will fall in the 0.60–0.75 range under the four-stream likelihood framework of Definition 4.6: suggesting but not decisively establishing polycentric emergence. *Falsified by*: a posterior $\Pr(r < 1) > 0.95$ (decisive polycentric evidence) or < 0.5 (diffusion preferred).

P9 The binding axiom for each technology should predict the most informative future evidence: improving evidence on the binding axiom should reduce posterior uncertainty in Σ more than improving evidence on non-binding axioms. *Falsified by*: a case where improving evidence on a non-binding axiom substantially shifts the posterior while improving evidence on the predicted binding axiom does not.

P10 Technologies in Regime III (high threshold, sparse contact) should show geographically confined traditions that do not spread even within the connected sub-network, producing a patchy distribution of the technology that is incompatible with both polycentric emergence and simple diffusion. *Falsified by*: a high-threshold technology showing spatially extensive spread in a demonstrably sparse-contact network.

P11 The transition from Regime I to Regime II (Note 2.1) should produce an observable shift from structural divergence to structural convergence in the material record of each technology, dateable to the period at which regional contact networks reached the technology-specific ρ^*. *Falsified by*: structural convergence beginning before the independently documented network densification, or persisting structural divergence after documented network densification.

P12 For horse domestication, improved palaeogenomic sampling of Central Asian and East Asian steppe populations at $c.\ 4{,}000$–$3{,}000$ BP should yield low inter-regional admixture (γ well below the contact threshold for the relevant date range) between Pontic-Caspian and Mongolian steppe populations, consistent with the *strong* Axiom IV profile. *Falsified by*: aDNA evidence for substantial Pontic-Caspian gene flow into Mongolian steppe populations before $3{,}500$ BP.

6.4 Conditions for Falsification of the Model as a Whole

Individual prediction failures are evidence against the model but do not by themselves refute it: any empirical theory can accommodate a small number of anomalous findings by invoking

local factors not captured in the model's parameters. The model as a whole is falsified under the following conditions:

Definition 6.1: Conditions for Whole-Model Falsification

The Network Horizon model in its general form (Chapters 1–2) is *falsified* if any of the following are demonstrated:

(F1) A technology with a *strong* or *moderately strong* overall axiomatic profile (predicted polycentric) is shown, under the full four-stream evidentiary framework, to have combined evidential weight *strongly* favouring diffusion. This would demonstrate that the axiomatic conditions are not jointly sufficient for polycentric emergence — that some axiom(s) are missing.

(F2) A technology with a *weak* overall axiomatic profile (predicted diffusion-dominant or very weakly polycentric) is shown to have four-stream combined evidence *decisively* favouring polycentric emergence. This would demonstrate that the axioms are not jointly necessary — that polycentric emergence can occur without satisfying the axiomatic conditions.

(F3) The axiomatic rank ordering across the four cases of Volumes I–IV is inverted by new evidence: specifically, if ceramics (predicted: highest polycentric proneness) is shown by better evidence to have stronger diffusion support than the wheel (predicted: lowest polycentric proneness), the rank ordering that the model generates is empirically incorrect and the axiomatic framework requires fundamental revision.

Conditions (F1)–(F3) define the scope of evidence that would require abandoning or substantially revising the model, rather than merely adjusting its parameter estimates.

6.5 Positioning the Network Horizon Model in the Theoretical Landscape

Case Study 6.1: The Network Horizon Model vs. Standard Cladistic Archaeology

Standard cladistic archaeology (O'Brien and Lyman 2000; Lycett 2009) treats artefact assemblages as analogues of biological organisms and applies phylogenetic methods to infer branching relationships between traditions. The implicit assumption is that technological change is predominantly descent-with-modification from a common ancestor, making diffusion the default model of inter-tradition relationship. The Network Horizon model explicitly challenges this assumption: it predicts that for technologies in Regime I, the correct phylogenetic structure is a set of independent lineages with no shared ancestor within the relevant chronological window, rather than a single tree. Cladistic methods applied to such assemblages will produce a tree regardless — the branching structure is an artefact of the method when the true history is polycentric. The Network Horizon model's contribution to this methodological debate is to specify *a priori* which technologies should and should not be analysed using tree-based phylogenetic methods, based on their axiomatic properties.

Case Study 6.2: The Network Horizon Model and Cultural Evolutionary Theory

Cultural evolutionary theory (Richerson and Boyd 2005; Henrich 2016) treats cultural transmission as analogous to biological inheritance and models the spread of cultural

variants through populations by social learning. Its core mechanism is transmission bias: certain variants are preferentially copied, producing cumulative cultural evolution without requiring individual innovation. The Network Horizon model is not in conflict with cultural evolutionary theory for the within-region phase of each tradition's development, where cumulative refinement through social learning is the primary mechanism. The conflict arises in the between-region comparison: cultural evolutionary theory's default assumption is that similar cultural variants in different regions are related by transmission, while the Network Horizon model's Axiom III predicts that similar functional solutions will arise independently wherever similar functional pressures apply. Distinguishing these two processes — convergent cultural evolution and parallel independent invention — requires exactly the multi-source evidentiary framework developed in Chapter 4, and cannot be resolved by cultural evolutionary theory's internal methods alone.

6.6 The Problem of Confirmation Bias in Prospective Prediction

Chapter 6's structural predictions (Finding 6.1) and diagnostic predictions (Finding 6.2) were formulated on the basis of the axiomatic framework developed in Chapter 2, which was itself constructed from four cases that all exhibit polycentric emergence. The risk of confirmation bias in this procedure — that the axiomatic framework is structured, consciously or otherwise, to generate favourable predictions for all cases with structural properties similar to the calibration set — requires explicit methodological engagement rather than simple denial.

The standard response to the confirmation bias objection is to note that prospective predictions are testable by future evidence that the model's author cannot control. This is correct as far as it goes, but it is an insufficient response in the absence of specific information about what kinds of evidence would count as falsifying the predictions. A prospective prediction that is compatible with all possible outcomes is not a prediction but a truism, and the Network Horizon model would be in this category if Finding 6.2's twelve diagnostic predictions were so vaguely stated that any reasonable outcome could be interpreted as consistent with them.

Definition 6.2: A Falsifiability Protocol for Network Horizon Predictions

For each diagnostic prediction P_j of Finding 6.2, a *falsifiability protocol* specifies: (i) the minimum sample size and evidential quality required for a test to count as a genuine test of P_j; (ii) the specific quantitative threshold that must be exceeded for the prediction to be considered confirmed; and (iii) the specific quantitative threshold whose violation counts as falsification. For example, for prediction P1 (zero-contact criterion for any technology with $\Sigma > 0.5$), the falsifiability protocol specifies: (i) INAA analysis of at least 30 artefacts per candidate region from the earliest securely dated context, using a reference database of at least 500 samples from the relevant geographic region; (ii) confirmation requires $\chi(R_i, R_j) < 0.05$ for all candidate pairs; (iii) falsification requires $\chi(R_i, R_j) > 0.15$ for any candidate pair with sufficient sample size. This protocol converts the qualitative prediction "zero cross-assignment rates expected" into a quantitative, decision-theoretic statement with specified evidential requirements.

Theorem 6.1: The Calibration Requirement for Prospective Prediction

Let $\hat{\Sigma}(\mathcal{T}^*)$ be the predicted composite polycentric score for a future technology $\mathcal{T}^*$ not in the calibration corpus. The calibration error in this prediction — the expected deviation of $\hat{\Sigma}$ from the true $\Sigma(\mathcal{T}^*)$ — depends on the size and diversity of the calibration

corpus used to set the model's parameters. For a calibration corpus of n cases uniformly distributed over the parameter space, the expected calibration error is proportional to $n^{-1/2}$, implying that a corpus of $n = 4$ cases yields calibration error approximately twice as large as a corpus of $n = 16$ cases. The current four-case corpus is therefore insufficient for precise point predictions of $\hat{\Sigma}$ for new cases, but it is sufficient for coarse-grained classification: the prediction that agriculture and textiles will show strong polycentric evidence while iron smelting will show weak polycentric evidence is robust to the calibration uncertainty even at $n = 4$, because the three predicted profiles are sufficiently separated (uniformly strong for agriculture, intermediate for textiles, weak for iron) that calibration errors of the magnitude implied by $n^{-1/2}$ scaling would not change the rank ordering.

6.7 The Iron Smelting Case in More Detail

Finding 6.1's prediction of weak polycentric evidence for iron smelting (overall axiomatic profile:
emphweak) deserves more detailed examination, because iron smelting is one of the cases most frequently discussed in the diffusionism versus polycentrism literature and because the prediction is non-trivial: it requires that the model correctly identify iron smelting as qualitatively different from ceramics despite sharing several structural features (widely distributed ore, high functional utility) that are superficially similar.

Case Study 6.3: Axiomatic Assessment of Iron Smelting

The axiomatic profile of iron smelting as a candidate for polycentric analysis:

Axiom I (Material Redundancy): Iron ores (magnetite, haematite, limonite, siderite) are among the most geographically widespread metallic ores on Earth, occurring in commercially significant deposits on every inhabited continent. Axiom I satisfaction:
emphstrong — comparable to ceramics.

Axiom II (Low Minimum Viable Threshold): This is the critical constraint. Smelting metallic iron from oxide ores requires sustained temperatures above $1{,}200^{\circ}$C in a strongly reducing carbon-monoxide atmosphere, the production of a bloom of metallic iron mixed with slag, and subsequent working of the bloom to separate iron from slag by hammering and reheating — a multi-step process that requires prior mastery of high-temperature kiln or bloomery construction and operation, empirical knowledge of reducing atmosphere management, and the skills to work a hot iron bloom without fracturing it. None of these prerequisites is achievable by a first-generation experimenter without prior pyrotechnological tradition; all require knowledge accumulated across many generations and transmitted within a practised tradition. Axiom II satisfaction: *weak* — markedly lower than copper smelting's *moderate* rating, and far below ceramics' *strong* rating.

Axiom III (Recurrent Functional Pressure): Iron's functional advantages over bronze and stone are substantial (harder cutting edge, more widely available raw material, lower production cost at scale) and would generate adoption pressure wherever iron was available and the skills to smelt it could be acquired. However, the functional pressure drives adoption and secondary spread rather than independent invention: communities that could obtain iron tools by trade had little incentive to invest in the considerable skill accumulation required for independent smelting. Axiom III satisfaction: *moderate–strong* (adoption pressure is strong but drives secondary spread rather than independent invention, which weakens the functional-pressure score for the polycentric case).

Axiom IV (Demographic Isolation): The earliest iron smelting attestations cluster in a relatively small geographic area (the Anatolian-Caucasian corridor and adjacent regions) within a chronological window ($c.$ 1200–1000 BCE) consistent with diffusion from a common tradition during the Bronze Age collapse. The contact network at this period was, despite the social disruptions of the late Bronze Age, substantially denser than the LGM and Chalcolithic networks relevant to ceramics and copper. Axiom IV satisfaction: emphmoderate (sits above the wheel on demographic isolation, but the dense Bronze Age contact network makes this the secondary binding constraint after Axiom II).

Axiom V (Functional Convergence Without Structural Homology): Iron-working traditions across the Old World show considerable structural similarity in bloomery design, carbon-atmosphere management, and bloom-working techniques — a pattern more consistent with shared tradition than independent development, although the degree of structural similarity is partially expected from the strongly convergent functional requirements of the technology. Axiom V satisfaction:
emphmoderate (structural similarity between Old World traditions is more consistent with shared tradition than independent development, though functional convergence partially explains the pattern).

Predicted overall axiomatic profile: *weak*, with Axiom II (minimum viable threshold) as the decisive binding constraint and Axiom IV (demographic isolation) as a secondary limiting factor. All other axes reach *strong* or *moderate–strong*, but the joint profile is below the wheel on the polycentric-proneness spectrum.

This is consistent with the assessment in Finding 6.1, which also placed iron smelting below the wheel in polycentric proneness; the revised profile confirms the prediction that iron smelting is the case *least* favourable to polycentric emergence across the technologies examined in this series, predicting weaker polycentric evidence for iron than for any of the four core cases.

Corollary 6.1: The Implication of the Iron Smelting Prediction for Sub-Saharan Africa

The iron smelting prediction of Case Study 6.3 applies specifically to Old World iron smelting from a common Anatolian-Caucasian tradition. It does not apply without modification to Sub-Saharan African iron smelting, which has a separate and earlier archaeological record that raises distinct questions about the independence of African iron metallurgy from Old World traditions. The earliest Sub-Saharan African iron smelting attestations (Central African Great Lakes region, $c.$ 800–700 BCE, and West African contexts possibly somewhat later) are geographically and chronologically separated from the Anatolian-Caucasian tradition in ways that might satisfy the chronological infeasibility criterion of Definition 4.2 for the specific region pair (Anatolia–Central Africa). The Network Horizon model's prediction for this specific pair depends on the assessment of Axiom IV for the Chalcolithic-to-Iron Age contact network between Anatolia and the Central African Great Lakes region — a contact network that is not well characterised by existing ancient DNA data. The Sub-Saharan African iron smelting question is therefore identified as a high-priority target for future aDNA research, with the potential to produce a Bayesian posterior $\Pr(r < 1)$ in the 0.6–0.85 range — suggestive polycentric evidence that would, if confirmed, require updating the iron smelting axiomatic profile in a direction that assigns higher weight to African metallurgical traditions' independence from Old World diffusion.

6.8 The Predictive Scope Beyond Technology

The Network Horizon model was developed specifically for prehistoric technologies, but its axiomatic structure is not in principle limited to material technologies. The five axioms characterise any domain in which complex functional solutions arise in multiple geographically separated populations responding to similar functional pressures with limited inter-regional contact: religious institutions, legal systems, writing systems, and mathematical concepts are all candidates for polycentric emergence analysis under the axiomatic framework.

Proposition 6.2: Extension of the Framework to Cultural Innovations

The axiomatic conditions for polycentric emergence (Axioms I–V, Chapter 2) can be restated for cultural innovations by substituting appropriate domain-specific primitives: for a cultural innovation $\mathcal{I}$ (a religious institution, legal norm, mathematical concept), Axiom I's material redundancy condition becomes a *cognitive resource redundancy* condition — the cognitive prerequisites for $\mathcal{I}$ are available in all candidate populations regardless of prior cultural contact; Axiom II's minimum viable threshold becomes the minimum population size and social-cognitive complexity sufficient to generate and maintain $\mathcal{I}$; Axiom III's functional pressure becomes the social or cognitive problem that $\mathcal{I}$ solves; and Axiom IV's demographic isolation condition remains structurally identical. Under this restatement, the framework predicts polycentric emergence for cultural innovations that satisfy the restated conditions, and it predicts single-origin diffusion for those that do not. The diagnostic tests differ (cultural phylogenetics and loan-concept analysis replace archaeometric provenance), but the inferential logic is structurally identical.

Note 6.1: The Limits of the Framework's Extension to Cultural Innovations

The extension proposed in Proposition 6.2 faces a significant obstacle not present in the material technology domain: the definition of the technological specification (Definition 2.1's criteria $\{c_1, \ldots, c_k\}$) is more straightforward for material technologies than for cultural innovations. A functional ceramic vessel satisfies a well-defined thermomechanical specification; a religious institution satisfies a social-functional specification that is substantially harder to define without circularity. The risk of defining the specification post-hoc in a way that makes convergence between two institutions appear stronger than it actually is — specifying the criteria of "a pantheon religion" in a way that covers both Mesopotamian and Mesoamerican religious systems — is considerably higher for cultural innovations than for material technologies. The framework's extension to cultural innovations is therefore potentially productive but requires a substantially higher standard of functional specification rigour than the material technology domain demands.

6.9 Failure Modes of the Prospective Prediction System

The prospective predictions of Findings 6.1 and 6.2 assume that the axiomatic scores estimated for new cases are unaffected by the investigator's prior knowledge of the case's outcome. In practice, this assumption is violated whenever the investigator conducting the axiomatic pre-assessment knows, from the secondary literature, whether polycentric emergence or diffusion is the currently accepted explanation for the technology under evaluation. The risk is that axiom scores are unconsciously inflated (for cases expected to show polycentric evidence) or deflated (for cases expected to show diffusion), producing a pre-assessment that appears to confirm the model but actually reflects the investigator's pre-existing beliefs about the case.

Definition 6.3: The Blind Pre-Assessment Protocol

To guard against outcome-aware axiom scoring, the Network Horizon framework recommends a *blind pre-assessment protocol* in which: (i) the axiomatic pre-assessment is conducted by an investigator with expertise in the relevant disciplinary domains (geology, experimental archaeology, palaeoecology, population genetics) but without knowledge of the specific chronological patterns under examination; (ii) the axiom scores are recorded and sealed before any examination of the first-attestation date distribution or existing interpretations of the convergence pattern; and (iii) the sealed scores are compared against the subsequent empirical analysis by a third party who did not participate in either the pre-assessment or the empirical analysis. This protocol cannot guarantee the elimination of unconscious bias — an investigator who knows that the framework predicts polycentric emergence for high-Σ cases may unconsciously inflate scores for cases believed to show polycentric evidence — but it substantially reduces the most overt forms of post-hoc confirmation and creates an auditable record of the pre-assessment process that can be examined by critics.

6.10 The Relationship Between Predictive Success and Theoretical Depth

A recurrent theme in the philosophy of science is the distinction between theories that make accurate predictions and theories that make accurate predictions for the right reasons — theories with genuine explanatory depth rather than merely empirical fit. The Network Horizon model's prospective predictions, if confirmed, would demonstrate predictive accuracy. The question of whether that accuracy reflects genuine explanatory depth is a separate, and harder, question.

Case Study 6.4: Deep vs. Superficial Predictive Success

Consider a simpler alternative to the Network Horizon model: a pure geographic distance rule, which predicts polycentric emergence for any two candidate regions separated by more than a threshold distance D_{min}, without any explicit axiomatisation of the structural conditions that make independent invention likely. Such a rule, calibrated to the four core cases, would also correctly classify agriculture and textiles as high-polycentric-probability technologies (both have candidate regions separated by trans-oceanic distances) and iron smelting as lower-polycentric-probability (its candidate regions are all within the Old World's connected landmass). The geographic distance rule has predictive success comparable to the Network Horizon model for the specific predictions of Finding 6.1, but for a theoretically impoverished reason: it does not explain why distance predicts polycentricity, while the Network Horizon model explains this through the mechanism of contact-density isolation (Axiom IV). The distinction matters for the range of cases to which the prediction generalises: the geographic distance rule fails for any technology with geographically separated candidate regions but above-threshold contact density (e.g., Bronze Age long-distance exchange networks), while the Network Horizon model correctly predicts diffusion for such cases because the isolation condition of Axiom IV explicitly incorporates contact density rather than merely geographic distance. Predictive success within the range of the calibration cases does not discriminate between theories at different levels of explanatory depth; extrapolation beyond the calibration range does.

6.11 A Worked Prospective Case: Early Pottery Firing Technique Variation

The prospective prediction methodology of this chapter has so far been illustrated primarily through the iron smelting case (treated as a calibration check rather than a true prospective test, since its outcome is already known) and brief discussion of candidate future cases. This section works through a genuinely prospective application in greater detail, illustrating the full prediction-and-evaluation cycle the methodology is designed to support.

Worked Example 6.4: Prospective Evaluation of Open-Firing Versus Kiln-Firing Transition Technology

Consider the transition from open-pit firing to enclosed kiln firing as a distinct sub-technology within the broader ceramics tradition, geographically and chronologically distributed across multiple Neolithic and Chalcolithic regions in a pattern not yet subjected to the full Network Horizon evidentiary protocol at the time of this volume's composition. Applying the axiomatic scoring framework prospectively: Axiom I (material redundancy) is plausibly satisfied, since the raw materials for kiln construction (clay, stone) are widely available; Axiom II (minimum viable threshold) requires assessment of the cognitive and engineering complexity of achieving sustained higher firing temperatures in an enclosed structure, a complexity that experimental archaeology suggests is non-trivial but achievable through incremental refinement of open-firing practice, suggesting a moderate-to-high satisfaction score; Axiom III (recurrent functional pressure) is satisfied insofar as higher, more controllable firing temperatures confer generalisable advantages (vessel durability, glaze and surface-finish possibilities) that would recur as a pressure across any ceramic-producing population; Axiom IV (demographic isolation) requires region-specific contact-density data not yet compiled with the precision needed for a confident score. On the basis of Axioms I–III alone, a preliminary composite score in the range $\Sigma \approx 0.5$–0.6 is estimated, placing this case in the steep intermediate range identified by Proposition 2.4, where the pending Axiom IV evidence is likely to be decisive.

Note 6.4: The Status of Worked Example 6.4 as a Genuine Prediction

Worked Example 6.4 is presented as a genuine prospective prediction in the sense defined by Chapter 6's falsifiability criteria: the preliminary composite score and its associated uncertainty are recorded here, prior to completion of the Axiom IV evidence-gathering, so that the eventual determination can be checked against this preliminary estimate. Should subsequent contact-network reconstruction for the relevant regions and periods establish either a high-density network (favouring diffusion-mediated spread of kiln technology from a single innovation locus) or a low-density network consistent with the demographic-isolation requirements of Axiom IV (favouring independent development of kiln firing in multiple regions under the shared functional pressure identified above), the result will constitute a genuine test of the framework's prospective predictive validity, in the spirit of the falsifiability criteria established earlier in this chapter. This is, to the authors' knowledge, the first case in the series presented explicitly in this pre-registered prospective format, and future volumes are intended to expand this practice.

6.12 Calibrating Prediction Confidence Against Track Record

As the prospective prediction programme accumulates a track record of resolved predictions, it becomes possible to ask a higher-order question: how well-calibrated are the framework's confidence levels, in the formal sense that predictions made with (say) 80% confidence should turn out to be correct approximately 80% of the time across a sufficiently large set of resolved cases.

Definition 6.5: Calibration Curve for Prospective Predictions

Given a set of resolved prospective predictions, each associated with a stated confidence level $c_i \in (0, 1)$ derived from the posterior probability $\Pr(\text{polycentric} \mid \mathbf{E})$ at the time of prediction, the empirical calibration curve plots, for each confidence bin (e.g., $[0.5, 0.6), [0.6, 0.7), ...$), the observed fraction of predictions in that bin that were subsequently confirmed correct against the bin's nominal confidence level. A well-calibrated prediction system produces a calibration curve close to the diagonal; systematic deviation above the diagonal indicates underconfidence (the framework is more often right than its stated confidence suggests), while deviation below indicates overconfidence.

Note 6.5: Insufficient Track Record for Current Calibration Assessment

At the time of this volume's composition, the number of genuinely prospective predictions made under the pre-registered protocol illustrated in Worked Example 6.4 is too small to construct a statistically meaningful calibration curve; a single resolved case, however its outcome turns out, cannot distinguish good calibration from chance agreement. This is an acknowledged limitation of the present state of the prospective prediction programme rather than a claim that calibration is unassessable in principle: as the programme expands to cover additional prospective cases (candidates include the kiln-firing transition of Worked Example 6.4, lithic microblade technology, and selected fibre-processing technologies not yet evaluated under the framework), the calibration assessment of Definition 6.5 becomes increasingly meaningful, and is intended as a standard component of future volumes' methodological apparatus.

6.13 Distinguishing Predictive Success from Retrodictive Fit

A persistent risk for any framework validated partly through retrospective case studies is that apparent predictive success may, on closer inspection, amount to little more than retrodictive fit dressed in predictive language. This section makes explicit the criteria the series uses to distinguish the two, building on the falsifiability discussion earlier in the chapter.

Proposition 6.5: Three Criteria for Genuine Predictive Status

A claim by the Network Horizon framework counts as a genuine prediction, rather than a retrodictive fit, only if three conditions are jointly satisfied: first, temporal priority, meaning the framework's classification is recorded before the relevant confirming or disconfirming evidence is examined (satisfied by the pre-registration protocol of Worked Example 6.4, not satisfied by the four core cases of Volumes I–IV, where axiom scoring and convergence-pattern evidence were necessarily considered together during the original research process); second, non-triviality, meaning the prediction is not so closely tied to features used in constructing the axioms that confirmation would be near-automatic (addressed by Corollary 2.4's parsimony constraint on axiom proliferation); and third,

risk, meaning the prediction could in principle have come out otherwise given the framework's own internal logic, ruling out predictions that are definitionally guaranteed to succeed regardless of the evidence. The four core cases of Volumes I–IV satisfy non-triviality and risk but not temporal priority, and are accordingly presented throughout this series as calibration cases establishing the framework's internal consistency rather than as predictive successes in the strict sense reserved for genuinely prospective cases like Worked Example 6.4.

Corollary 6.4: The Limited but Real Evidential Value of Calibration Cases

Calibration cases that satisfy non-triviality and risk, even without temporal priority, retain genuine evidential value distinct from pure retrodictive fit: a framework whose axioms could not, even in principle, have classified the four core cases as anything other than polycentric (failing the risk criterion) would provide no evidential support regardless of temporal priority, whereas a framework whose axioms genuinely could have classified one or more of the four cases as diffusion-favouring, and did not, provides some evidential support for the framework's discriminating power even in the absence of strict prospective status. The iron-smelting negative control of Worked Example 1.3 strengthens this evidential value further by demonstrating that the same axiomatic apparatus, applied to a case independently known to be diffusion, correctly produces a diffusion-favouring classification, addressing the risk criterion directly for at least one case in the broader evidential landscape considered by the series.

6.14 A Second Worked Prospective Case: Textile Loom Technology

To broaden the prospective prediction programme beyond the single worked case of Worked Example 6.4, this section develops a second prospective case, chosen specifically because it differs structurally from the kiln-firing transition case in ways that test the framework's applicability across a wider range of technological types.

Worked Example 6.5: Prospective Evaluation of Warp-Weighted Loom Technology

Warp-weighted loom technology, attested archaeologically through the characteristic loom weights it leaves in the material record, presents a contrasting profile to the ceramic and metallurgical cases examined elsewhere in the series: its functional pressure (efficient production of wider and more structurally consistent woven textiles than achievable with simpler back-strap or ground looms) is plausibly recurrent across textile-producing populations, satisfying Axiom III; its threshold accessibility (Axiom II) is moderate, since the loom-weight principle, once conceived, requires only modest material and engineering sophistication, but conceiving the principle itself — using gravity-tensioned weights to maintain consistent warp tension across a wide weaving frame — represents a non-trivial cognitive step without obvious precedent in simpler textile technologies; material redundancy (Axiom I) is satisfied broadly, since loom weights can be fashioned from fired clay, stone, or other locally available dense materials, imposing few geographic constraints; demographic isolation (Axiom IV) requires region-specific contact-network analysis not yet compiled for the relevant Bronze Age and Chalcolithic contexts in which warp-weighted looms are first attested. On the basis of Axioms I–III, a preliminary composite score in the range $\Sigma \approx 0.45$–0.55 is estimated, again placing this case in the steep intermediate range of Proposition 2.4, though somewhat lower than the kiln-firing case owing to the greater cognitive non-triviality assessed for Axiom II.

Note 6.6: Why Textile Technology Was Not Included in the Core Calibration Corpus

Warp-weighted loom technology was considered, alongside basketry and bow-and-arrow technology (Note 1.6), as a candidate for the core calibration corpus, but excluded primarily on Axiom IV evidential-completeness grounds: the demographic and contact-network reconstruction needed to assess regional isolation with confidence is not yet available at sufficient resolution for the Bronze Age and late Chalcolithic contexts in which warp-weighted looms are first well attested, a period for which the ancient-DNA evidentiary stream, while improving rapidly, remains less comprehensive than for the earlier Neolithic contexts of the four core cases. This evidentiary gap is expected to narrow as ancient-DNA sampling density for Bronze Age contexts continues to improve, and warp-weighted loom technology is identified here as a strong candidate for eventual promotion from the prospective prediction set to the core calibration corpus once Axiom IV evidence reaches adequate resolution.

6.15 Comparing the Two Worked Prospective Cases: A Methodological Reflection

The two worked prospective cases developed in this chapter — kiln-firing transition technology and warp-weighted loom technology — share the structural feature of falling in the steep intermediate range of preliminary composite scores, with Axiom IV evidence pending in both cases. This pattern, while based on only two cases, is worth flagging as a possible general feature of technologies that fail the evidential-completeness criterion for the core corpus.

Proposition 6.6: A Conjecture on Why Excluded Technologies Cluster in the Intermediate Range

Technologies excluded from the core calibration corpus on Axiom IV evidential-completeness grounds are disproportionately likely, on the present limited evidence, to show preliminary composite scores (computed from Axioms I–III alone) in the intermediate range rather than at the extremes, because Axioms I–III depend on geological, experimental-archaeological, and ecological evidence that is generally available even for technologies whose contact-network reconstruction (Axiom IV) remains incomplete, while the steep-intermediate-range pattern itself may simply reflect that Axioms I–III alone are less discriminating than the full five-axiom combination, since they omit the contact-density information that, per the dominant-stream analysis of Worked Example 3.5, often supplies much of the discriminating power for borderline cases. If this conjecture holds more generally, it suggests that the prospective prediction programme should expect a persistent population of intermediate-confidence cases awaiting Axiom IV resolution, rather than a population already skewed toward confident classification before contact-network evidence becomes available — a pattern with direct implications for how research effort should be allocated, since it implies that closing the Axiom IV evidence gap, rather than further refining Axioms I–III evidence collection, offers the greatest expected gain in classification confidence across the prospective case population as a whole.

Corollary 6.5: A Research Prioritisation Recommendation

Given Proposition 6.6's conjecture, the highest-value direction for expanding the evidentiary basis of future Network Horizon applications is targeted ancient-DNA and exchange-network sampling specifically aimed at resolving Axiom IV for currently pending prospective cases, rather than further refinement of the geological, experimental-

archaeological, or ecological evidence underlying Axioms I–III, where the existing evidentiary base is already comparatively strong for most candidate technologies. This recommendation is offered as a practical guide for future fieldwork and sampling-design decisions by researchers seeking to extend the calibration corpus or resolve pending prospective cases, complementing the broader geographic and temporal expansion priorities identified in the concluding chapter's discussion of the calibration corpus's current limitations.

6.16 A Third Worked Prospective Case: Basketry and Plant-Fibre Technology

The two worked prospective cases developed elsewhere in this chapter and its companion supplementary sections — textile loom technology and early pottery firing technique variation — illustrate the framework's prospective application to technologies of moderate and well-attested archaeological visibility. Basketry and plant-fibre technology presents a methodologically distinct prospective case precisely because its archaeological visibility is exceptionally poor: plant fibre artefacts rarely survive except under exceptional preservation conditions (waterlogged, frozen, or extremely arid contexts), meaning the direct material evidence bearing on Axioms I, II, and V is far sparser than for any of the four core cases or the other prospective examples, while indirect evidence (impressions in clay, residue analysis, and use-wear on associated tools) is comparatively more available. This case therefore tests the framework's behaviour under conditions of structurally poor — rather than merely currently incomplete — direct evidence.

Definition 6.6: Indirect Evidential Proxies for Poorly Preserved Technologies

For a technology $\mathcal{T}$ whose direct material remains are subject to severe taphonomic loss, an *indirect evidential proxy* is a class of more durable archaeological correlate whose presence implies, with some specified reliability, the prior existence of $\mathcal{T}$ even in the absence of surviving direct remains. For basketry and plant-fibre technology, the principal indirect proxies are: cord-marked and basket-impressed ceramic surfaces, which preserve negative impressions of plant-fibre objects pressed into wet clay before firing; bone and antler tools bearing use-wear patterns consistent with fibre-processing activities (retting, splitting, twining); and, where preservation permits, phytolith and starch-grain residues on associated tool surfaces indicating the specific plant taxa processed. Each proxy carries a different reliability for inferring the presence of basketry specifically as opposed to other plant-fibre uses (cordage, matting, textiles), and a rigorous prospective application of the framework to this case requires explicit reliability weighting of each proxy type rather than treating indirect evidence as uniformly informative.

Case Study 6.5: The Convergent Cord-Marking Evidence Across Early Holocene Ceramic Traditions

Cord-marked and basket-impressed pottery is documented across an exceptionally broad geographic range in early Holocene contexts: Jōmon ceramic traditions in Japan from approximately $10{,}500$ BP, early ceramic traditions in the Russian Far East and Northeast Asia from a comparable period, early Holocene pottery in the southeastern United States from approximately $4{,}500$ BP, and basket-impressed wares in early West African ceramic traditions. This geographic and temporal breadth, considered as Axiom V evidence (functional convergence without structural homology, here applied to the up-

stream plant-fibre technology rather than the ceramic technology itself), is suggestive of widespread, plausibly independent basketry traditions underlying the impressed-ware evidence, but the inferential chain is longer and weaker than for any of the four core cases: cord-marking demonstrates the existence of cordage and basketry at the time and place of the ceramic find, but does not by itself demonstrate that the basketry tradition itself arose independently in each region rather than diffusing as a component of a broader technological and cultural package that included ceramic production. Disentangling these possibilities requires evidence beyond the cord-marking pattern itself, which the current archaeological record only partially provides.

Theorem 6.7: Indirect-Proxy Evidence Yields Wider Composite Score Intervals

For a technology assessed primarily through indirect proxies (Definition 6.6) rather than direct material remains, the composite score interval $[\hat{\Sigma}_{\text{min}}, \hat{\Sigma}_{\text{max}}]$ computed under the missing-evidence propagation rule of Definition 2.7 is systematically wider than for technologies with comparable direct evidential access, because each indirect proxy carries an additional layer of inferential uncertainty (the proxy-to-technology inference itself) beyond the ordinary evidential uncertainty captured in the four core cases' scoring. For basketry and plant-fibre technology, applying the axiom-scoring protocol to the available indirect proxies yields a preliminary estimate of $\hat{\Sigma}(\text{basketry}) \approx 0.55$–$0.75$, a wider interval than any of the four core cases or the other two prospective cases discussed in this chapter, reflecting genuine uncertainty rather than merely insufficiently careful estimation. This wide interval is itself an informative output of the framework: it identifies basketry and plant-fibre technology as a case where the qualitative classification (more likely than not polycentric, given that the interval lies entirely above $\Sigma = 0.5$) is reasonably confident even though the precise composite score is not, a distinction the point-estimate-only reporting convention criticised in the discussion of Definition 2.7 would obscure.

Corollary 6.6: Basketry Illustrates the Framework's Behaviour at the Edge of Its Evidentiary Applicability

The basketry case, considered alongside the textile loom and pottery-firing-variation cases discussed elsewhere in this chapter, demonstrates that the Network Horizon framework's applicability does not require direct material evidence of the quality available for the four core cases; it requires only that the indirect proxy evidence, however attenuated, be sufficient to populate the axiom-satisfaction scores with non-trivial confidence intervals rather than complete ignorance. Cases for which even this weaker requirement cannot be met — where no proxy evidence of any kind bears on one or more axioms — fall outside the framework's current applicability and should be reported as such (using the explicit NA notation of Definition 2.7) rather than assigned arbitrary midpoint scores. Basketry sits close to this boundary without crossing it: the cord-marking and use-wear proxies provide genuine, if attenuated, evidential purchase on Axioms I, II, and V, while Axiom IV (demographic isolation) remains substantially under-evidenced for this case pending the kind of targeted ancient-DNA sampling Proposition 6.6 elsewhere identifies as the highest-priority direction for extending the framework's prospective applications.

Note 6.7: Why This Case Was Not Promoted to the Calibration Corpus

Despite the suggestive convergent evidence discussed in Case Study 6.5, basketry and plant-fibre technology has not been promoted from the prospective prediction set to the calibration corpus proper, in keeping with the evidential-completeness criterion

of Proposition 1.3: the requirement that at least two of the four evidence streams provide discriminating evidence (posterior likelihood ratio greater than 3 in either direction) is not currently met, because the genetic and linguistic streams remain essentially silent on this case (no systematic ancient-DNA sampling programme has targeted populations specifically for fibre-technology-relevant questions, and no glottochronological reconstruction of basketry-related vocabulary across the relevant language families has been undertaken). The chronological and archaeometric streams alone, while suggestive, do not meet the two-stream minimum, and the case is accordingly reported here as illustrative of the framework's prospective range rather than as a fifth confirmed instance of polycentric emergence alongside the four core cases.

6.17 A Fourth Falsification Class: Predictions About Future Archaeological Discoveries

The twelve diagnostic predictions of Finding 6.2, together with the falsifiability protocols of Definition 6.2, are formulated primarily as predictions about how existing or readily obtainable evidence should appear once examined through the four-stream framework. A further class of falsifiable prediction, distinct in character, concerns specific, currently unexcavated or unsequenced material that the framework predicts should exist and should, when examined, exhibit specific properties — a stronger and more genuinely prospective form of falsifiability than re-examination of already-available evidence.

Definition 6.7: Discovery-Contingent Predictions

A *discovery-contingent prediction* specifies, in advance of any relevant excavation, sequencing, or survey having been conducted, what a specific not-yet-obtained body of evidence should show if the framework's classification of a given case is correct, in a form precise enough that the prediction could be falsified by the actual discovery once made. Unlike the predictions of Finding 6.2, which can in principle be tested against currently available but not-yet-analysed evidence, discovery-contingent predictions require evidence that does not yet exist in analysed form and may not exist at all, making them the most stringent and the most genuinely prospective class of falsifiable claim the framework offers.

Case Study 6.6: A Discovery-Contingent Prediction for the Wheel Case

The framework predicts, contingent on the wheel case's classification remaining ambiguous under the modality-corrected diffusion-velocity analysis of Theorem 4.6, that systematic ancient-DNA and isotopic strontium sampling of human and faunal remains from Pontic-Caspian steppe sites dated to the relevant window (c. 4,000–3,500 BCE) should reveal mobility and exchange-network signatures consistent with the riverine-corridor contact intensity estimated in Worked Example 4.8, specifically: isotopic evidence of non-local origin (indicating long-distance movement of individuals, consistent with active exchange networks) in a proportion of sampled individuals from riverine-corridor-adjacent sites exceeding the proportion from sites without riverine access, by a margin large enough to be statistically distinguishable given plausible sample sizes (on the order of 30–50 individuals per site category). This prediction is falsifiable by future isotopic survey work and would, if falsified (finding no differential mobility signature between riverine-adjacent and riverine-distant sites), substantially weaken the riverine-corridor correction's case for shifting the wheel classification toward diffusion feasibility, strength-

ening instead the independent-invention reading of the wheel case.

Proposition 6.7: Discovery-Contingent Predictions Carry Higher Evidential Value Than Re-Analysis Predictions

A discovery-contingent prediction (Definition 6.7) that is subsequently confirmed by genuinely new evidence carries higher evidential weight for the framework's validity than confirmation of a prediction tested against already-available evidence (the predictions of Finding 6.2), because the discovery-contingent prediction was made without any possibility that the prediction was shaped, even unconsciously, by foreknowledge of the specific result, satisfying the temporal-priority consideration discussed in connection with the risk criterion of Worked Example 1.3 more fully than re-analysis predictions can. The framework's overall falsifiability case would be substantially strengthened by accumulating a portfolio of discovery-contingent predictions of the kind illustrated in Case Study 6.6, across all four core cases and the prospective cases of this chapter, as a priority for the research programme this volume opens rather than closes.

Note 6.8: Why Only One Discovery-Contingent Prediction Is Offered in Detail Here

The single detailed discovery-contingent prediction offered in Case Study 6.6, rather than a comprehensive portfolio across all four core cases, reflects a practical constraint rather than a principled limitation: formulating a rigorous discovery-contingent prediction requires detailed knowledge of which specific sites are realistic candidates for future sampling, what sampling protocols are methodologically appropriate, and what statistical power is achievable given realistic sample-size constraints, a level of site-specific and methodological detail that exceeds what could be responsibly developed for all four cases within the scope of this volume. The wheel case was selected for detailed treatment here because, as established repeatedly throughout this volume, it is the case whose classification is most genuinely open and therefore the case for which a discovery-contingent prediction carries the most diagnostic value; future companion technical materials, prepared in collaboration with archaeologists possessing the relevant site-specific expertise, are the appropriate venue for extending this treatment to the other three core cases.

Corollary 6.8: What the Formal Apparatus of This Chapter Has and Has Not Done

It is worth separating, plainly and at the close of the chapter that contains the model's most elaborate formal machinery, what the axiomatic and Bayesian apparatus of Chapters 2–6 actually accomplishes from what its notational density might suggest it accomplishes. The apparatus has done three things: it has made the four core cases' qualitative classification (Σ uniformly strong for ceramics, intermediate for the wheel, and so forth) explicit and internally consistent rather than implicit and case-specific, as it was distributed across Volumes I–IV; it has generated a small number of falsifiable, quantitatively specified predictions for cases outside the calibration set (textiles, iron smelting, and the discovery-contingent prediction of Case Study 6.6); and it has made the conditions for its own refutation (Definition 6.1's F1–F3) explicit in advance rather than reconstructable only after the fact. It has not done, and should not be read as having done, two further things that formal notation can sometimes be mistaken for providing: it has not converted a four-case qualitative judgment into a statistically powered quantitative theory — Proposition 7.1's calibration-error analysis applies in full force to every score and posterior odds figure in this volume — and it has not supplied independent confirmation beyond what the underlying archaeological, genetic, isotopic, and linguistic evidence itself provides. The axioms

organise the evidence; they are not a second, independent source of evidence alongside it. A reader who finds the apparatus persuasive should locate that persuasiveness in the calibration-set case studies of Chapter 3 and the prospective predictions of this chapter, not in the existence of the formal structure as such.

Chapter 7

Limits, Open Questions, and the Horizon Ahead

7.1 What This Series Has and Has Not Established

Four volumes of multi-source archaeological, archaeometric, genetic, and linguistic analysis have established, to a high level of evidential confidence, that ceramic technology, horse domestication, copper metallurgy, and wheeled transport each satisfy the criterion for polycentric emergence set out in Definition 1.1: the available evidence supports the polycentric model over the diffusion model, with posterior odds ranging from 7.1 (wheel) to more than 10,000 (ceramics) under the Bayesian integration framework of Definition 4.6.

This is a substantial empirical result. It is not, however, a general theory of prehistoric technological change. Four cases, however carefully documented, do not establish that polycentric emergence is the *dominant* mode of prehistoric technological innovation — only that it occurs, and occurs systematically for a class of technologies satisfying the Network Horizon axiomatic conditions. What remains uncertain is the size of that class relative to the full population of prehistoric technologies, most of which have not been subjected to the same quality of multi-source analysis. The series documents the existence of a phenomenon; it does not yet establish its frequency.

Note 7.0: The Epistemological Legitimacy of Boundary-Case Analysis — The Wheel as a Test of the Framework

A recurrent question in the reception of Volume II of this series — which develops the wheel case in a form that includes sustained engagement with a specific monocentric diffusionist hypothesis (Haarmann 2016) and analyses the evidence for and against that hypothesis across four independent analytical motors (taphonomic, archaeogenomic, structural-mechanical, and historical-linguistic) — concerns the epistemological standing of a boundary-case analysis: is a case that the framework itself classifies as near the Regime I/II boundary (posterior odds $\Lambda = 7.1$, credible interval straddling $r = 1$) a legitimate target for extended critical treatment of a monocentric alternative?

The answer is affirmative, and the present framework's theoretical structure establishes why. The boundary position of the wheel case is not a weakness of the analysis but its most distinctive and most productive feature. Precisely because the wheel sits near $r = 1$, it is the case for which the four evidentiary streams can most directly constrain a specific monocentric alternative without producing a trivially decisive result in either direction. The criteria for adequate critical treatment of a monocentric hypothesis do not require that the polycentric alternative be decisively established; they require that

the monocentric hypothesis be subjected to the same evidentiary standards as the polycentric alternative, and that the analysis identify precisely where the current evidence is, and is not, sufficient to discriminate between them.

Volume II's treatment of Haarmann (2016) satisfies these criteria in the following respects: (a) the taphonomic motor establishes that the geographic concentration of surviving wheel-related wood finds reflects differential preservation coefficients of three to four orders of magnitude between anaerobic lacustrine environments and open-steppe pedology, not a geographic concentration of original production activity, directly addressing the error of treating the archaeological distribution of preserved material as a map of invention; (b) the archaeogenomic motor establishes that the biological contact between Trypillia and steppe populations was already at 15–25% admixture by 3800–3600 BCE — substantially pre-dating the proposed bilingual co-invention — demolishing the premise of biologically isolated cultures meeting at a single contact interface; (c) the structural-mechanical motor establishes a kinematic discontinuity of five to seven orders of magnitude between the tournette's operating regime (low-load, high-speed axial rotation) and the cart axle's regime (high-load, low-speed flexural-torsional combined stress), ruling out evolutionary derivation of the latter from the former by technological transmission without co-invention of load-path geometry, lubrication chemistry, and arsenical copper tooling; and (d) the historical-linguistic motor establishes, through the three-criterion substrate-evidence protocol of Section 8.2.3 applied to all eleven candidate terms in Haarmann's proposed substrate corpus, a score of 0/11 qualifying terms, formally excluding the proposed paleo-European lexical contribution from the PIE wheel vocabulary at the standard of evidence required by Thomason & Kaufman (1988) for demonstrated substrate contact.

These four results are not claims that polycentric emergence is decisive for the wheel; they are claims that the specific monocentric co-invention scenario of Haarmann (2016) is not supported by the evidence on four independently falsifying grounds, while the Network Horizon boundary-polycentric position remains consistent with the evidence on all four grounds. This is the epistemologically correct outcome for a boundary case: the monocentric hypothesis is excluded for specific, articulable reasons without the polycentric alternative being positively demonstrated beyond reasonable doubt. The boundary-case treatment of Volume II therefore occupies exactly the evidential position that the present framework's theoretical structure predicts and requires for a case at $r \approx 1$ — and it is in this sense that Volume II's extended critical analysis of a specific monocentric hypothesis constitutes a methodologically legitimate contribution to the series rather than an overreach of the framework's evidential scope.

Note 7.1: The Selection Bias Problem, Including Its Geographic and Temporal Dimension

The four cases examined in this series were not selected from a random sample of prehistoric technologies, and they share two compounding selection effects worth stating together rather than separately. First, they were selected because they are among the most archaeologically and linguistically tractable cases: technologies with rich material records, well-documented earliest attestations, strong palaeogenomic coverage, and reconstructable proto-language vocabulary domains. Technologies whose polycentric emergence left weak material, genetic, or linguistic traces may be systematically underrepresented in a corpus assembled on tractability grounds, and the direction of the resulting bias on the true frequency of polycentric emergence is not determinable from within this series. Second, and compounding the first effect, all four cases are drawn from the Old World and its immediate Eurasian-adjacent New World periphery, and all fall within the

Neolithic-to-Chalcolithic window; the series cannot rule out that some unidentified feature specific to this geographic and temporal window, rather than the genuinely general structural conditions the axioms aim to capture, is responsible for the pattern of results obtained. Both effects point to the same remedy: extending the calibration corpus to technologies from other world regions (sub-Saharan African metallurgical traditions, Pacific maritime technologies, Andean and Amazonian technological complexes) and other time periods, selected independently of tractability, is identified as the single highest-priority direction for future work building on this series.

7.2 The Inference from Four Cases to a General Pattern

The inductive step from four documented cases to the general claim that polycentric emergence is a structurally predictable outcome for technologies satisfying the axiomatic conditions is the most epistemically exposed move in the Network Horizon framework. It is supported by the retroactive confirmation from agriculture (Proposition 6.1) and by the internal consistency of the axiomatic structure across all four cases, but it rests on a sample size that limits statistical confidence.

Proposition 7.1: The Minimum Sample Size for Robust Generalisation

For the composite polycentric score $\Sigma(\mathcal{T})$ to function as a reliable predictor of empirical polycentric evidence across technologies not included in the calibration set, the model's parameters must be estimated from a sufficiently large and diverse sample. Under a simple bias-variance decomposition, the prediction error for a new technology $\mathcal{T}^*$ is:

$$\mathbb{E}[(\hat{\Lambda}(\mathcal{T}^*) - \Lambda_{\text{true}}(\mathcal{T}^*))^2] = \text{Bias}^2 + \text{Variance},$$

where the variance term scales inversely with the number of calibration cases. With $n = 4$ cases, the variance contribution to prediction error is non-negligible: the ± 1 standard error interval on the predicted composite score for a new technology spans approximately ± 0.15 on a $[0, 1]$ scale, which is sufficient to shift a borderline prediction (e.g., $\hat{\Sigma} = 0.40$) from one side of the Regime I/II boundary to the other. Reducing this uncertainty to the level where borderline predictions are reliable requires a calibration set of $n \geq 10$–15 cases, spanning the full range of the $(\theta/\theta_F, \bar{\lambda}/\rho^*)$ parameter space. Expanding the calibration set is therefore the primary methodological priority for the next phase of the research programme.

7.3 The Three Most Productive Open Questions

Of the many questions the Network Horizon framework raises without fully answering, three are singled out as the most productive for future research: the cognitive mechanism question, the failure-mode question, and the network densification question.

Open Question 7.1: The Cognitive Mechanism of Independent Invention

The Network Horizon model treats the invention event $\iota(R, \mathcal{T}, t)$ as a primitive (Definition 2.1): it specifies the structural conditions under which independent invention events are expected to occur but says nothing about the cognitive process by which any specific invention event unfolds. The question of whether independent inventors in South China and the Sahara solved the problem of fired clay vessels through similar cogni-

tive sequences — similar trial-and-error trajectories, similar analogical reasoning from prior container technologies, similar social contexts of technological discovery — is not addressed by the framework. Nor, with the available evidence, could it be. But it is a question of genuine theoretical importance for understanding whether polycentric emergence reflects deep regularities in human technical cognition or merely the recurrence of favourable material conditions. Cross-cultural experimental studies of ceramic making in populations with no prior exposure to fired ceramics (analogously to the cross-cultural studies of tool-making in non-human primates) could begin to address this question at the micro-scale, though the methodological challenges are substantial.

Open Question 7.2: The Failure-Mode Distribution

The Network Horizon model predicts that polycentric emergence will occur with high probability when the five axiomatic conditions are jointly satisfied. But it does not — and with four calibration cases cannot — estimate the failure-mode distribution: the proportion of cases where all five conditions are satisfied but independent invention nonetheless fails to occur within the expected chronological window. Such failure modes exist: there must have been many prehistoric communities satisfying all the axiomatic conditions for ceramic invention whose communities disappeared, whose clay deposits were unsuitable in ways not captured by Axiom I, or whose functional pressures were met by non-ceramic alternatives before the ceramic invention threshold was crossed. The failure modes are archaeologically invisible — they leave no material trace — and therefore cannot be counted from the surviving evidence. Without knowing the failure-mode rate, the model cannot be used to make probabilistic predictions about invention rates: it predicts that polycentric emergence is *possible and likely* given the axiomatic conditions, but not how likely in any quantitative sense.

Open Question 7.3: The Network Densification Trajectory and the Regime I/II Transition

Note 2.1 predicted that technologies beginning in Regime I would transition to Regime II as regional contact networks densified, producing a datable shift from structural divergence to structural convergence in the material record. Identifying and dating this transition for each of the four technologies examined in the series would provide an independent test of the model's contact-density mechanism: if the transition dates correlate with independently documented network densification events (the Yamnaya expansion for horse domestication, the Bronze Age exchange network for copper metallurgy, the Silk Road for ceramics in its later phases), the model's causal story — that it is the network density crossing ρ^* that drives convergence, not other factors — is corroborated. Documenting these transitions requires diachronic comparison of structural divergence measures across the full temporal span of each tradition, from the polycentric emergence phase through the convergence phase, at a level of chronological and geographic resolution that the current evidence base only partially supports. This is a tractable but labour-intensive research programme that spans several subdisciplines and would require a dedicated collaborative project beyond the scope of this series.

7.4 A Note on the Significance of Polycentric Emergence for Understanding Human Cognition

The most far-reaching implication of the Network Horizon framework is one that the framework itself, as a theory of technological emergence conditions, does not directly address: what

the recurrence of polycentric invention implies about human cognitive universals.

If the same technologies arise independently in multiple isolated populations, the obvious explanation is that human cognitive architecture is sufficiently uniform across populations to converge on similar solutions to similar problems when the material conditions permit. This is not a trivial claim. It is compatible with a strong version of the psychic unity of humankind — the proposition that the cognitive capacities relevant to technical problem-solving are invariant across human populations — and it is inconsistent with a strong diffusionist model of cultural evolution that treats technological knowledge as a scarce, locally-specific resource transmissible only by direct contact.

The Network Horizon framework does not establish cognitive uniformity; it establishes that the *outcomes* of technological problem-solving are convergent across isolated populations given similar material and functional conditions. Whether this outcome-convergence reflects cognitive uniformity, or merely the strong constraints imposed by the physics and chemistry of the relevant materials (clay, metal, fibre, horses), is a question that the archaeological evidence cannot resolve. But it is a question that the archaeological evidence, for the first time and through the evidential architecture documented in this series, has made precisely askable.

Note 7.2: The Network Horizon as a Research Programme, Not a Conclusion

It is tempting to read the Network Horizon series as arriving at a conclusion: polycentric emergence is real, widespread, and structurally predictable. That reading is not wrong, but it is incomplete. The more accurate characterisation is that the series has established the *existence proof* for a research programme: it has demonstrated that the multi-source evidential architecture of Definition 4.6, applied systematically to a technology with the right axiomatic properties, can resolve the Diffusion Boundary Problem with high posterior confidence. That demonstration opens, rather than closes, the field. The agenda it sets — expanding the calibration set, documenting the Regime I/II transition, estimating failure-mode rates, extending the framework to biological and cognitive technologies — is larger than what four volumes can accomplish. The horizon, as the series title implies, is always ahead.

7.5 Closing Remarks

The four volumes of this series were motivated by a recurring observation in the archaeological literature: that the standard diffusionist narrative, applied reflexively to any technology appearing in more than one region, was producing interpretations that the evidence, rigorously examined, did not support. The ceramics case was the most egregious example: a technology with first attestations spanning twelve millennia and four continents, explained by diffusion from a single origin despite the complete absence of any plausible diffusion mechanism that could bridge those spans.

The response developed across the series was not simply to assert polycentric emergence as an alternative narrative but to build the evidential and formal apparatus necessary to *decide* between the two models in a principled way. That apparatus — the four-stream Bayesian framework, the axiomatic model, the composite scoring system, the diagnostic predictions — is the series' principal theoretical contribution. The empirical findings about ceramics, horses, copper, and the wheel are important, but they are important primarily because they demonstrate that the apparatus works: that it generates determinate answers to questions that were previously answered by assumption.

The diffusionist default is not merely an interpretive convention; it reflects a genuine prior probability argument about the frequency of independent invention. The Network Horizon

framework does not deny that prior: it provides the methodological tools to update it appropriately in the light of evidence. Whether the posterior, once computed, favours diffusion or polycentric emergence depends on the evidence and the technology. For the four cases examined here, and for the wider class of technologies sharing their axiomatic properties, the evidence consistently updates the prior toward polycentric emergence. But that update is not a doctrine; it is a finding, subject to revision as the evidence improves and the calibration set expands.

The horizon moves as we approach it. That is not a limitation; it is the condition of empirical inquiry.

7.6 The Boundary Between This Series and Its Successors

The Network Horizon series occupies a specific temporal, methodological, and evidential niche: it applies the full multi-source evidential architecture of Chapter 4 to four Neolithic and Chalcolithic technologies for which that architecture is, at this moment, adequately equipped. It does not attempt to extend the analysis beyond what the current state of the evidentiary toolkit can support with the required rigour. Understanding what lies on the other side of this boundary — what questions are now askable that were not askable before the series — is essential for directing the research programme the series opens.

The boundary is defined by three converging limitations: the ancient DNA coverage of prehistoric populations, the INAA and lead isotope reference databases for archaeological materials, and the glottochronological calibration horizon. Each limitation defines a class of questions that the current framework cannot address and that advances in the relevant field would open.

Proposition 7.2: The Next Three Evidential Advances That Would Extend the Framework

The following three advances in the evidentiary sciences would, if achieved, extend the Network Horizon framework to a substantially expanded set of questions:

(i) Ancient DNA from tropical and subtropical contexts: The current aDNA coverage of prehistoric populations is heavily biased toward cold-climate contexts (steppe, northern Europe, the Levant) where DNA preservation is good. Extending systematic sampling to the tropical and subtropical contexts where ceramics (Sahara, Amazonia) and agricultural polycentrism (New Guinea, West Africa, Amazonia again) are argued to have occurred would transform the genetic stream of the diagnostic framework from a two-stream limitation to a four-stream instrument for those cases.

(ii) Machine-learning-assisted INAA source discrimination at multi-continental scale: The current INAA framework for ceramic fabric discrimination uses reference databases compiled from specific geographic regions and requires a priori assignment of samples to candidate source regions. An automated multi-continental source discrimination system, trained on the full global INAA database and capable of assigning ceramic samples to source regions without prior geographic constraint, would both expand the geographic scope of the material provenance stream and reduce the investigator-degree-of-freedom problem in source assignment.

(iii) Extension of the glottochronological calibration horizon: The current 6,000–8,000 year calibration horizon for Bayesian phylogenetic glottochronology reflects the limit at which internal calibration points with independently dated archaeological correlates can be identified in most language families. Improved palaeoanthropological and ancient DNA dating of language family dispersal events could push this horizon to 10,000–12,000 years, bringing the earliest ceramic attestations (at $c.$ 20,000 BP in South China)

within range of indirect linguistic constraint.

7.7 The Epistemological Status of the Network Horizon's Central Claim

The central claim of the Network Horizon series is that polycentric emergence — independent, structurally predicted re-invention of the same technology in multiple geographically separated contexts — is not an exceptional event requiring special explanation but the structurally expected outcome for a class of technologies defined by the axioms of Chapter 2. This claim occupies an intermediate position in the epistemological landscape between empirical finding and theoretical postulate: it is more than a description of the four cases (which would be a purely empirical statement) and less than a universal law of technological innovation (which would require far more cases than the series provides).

Definition 7.1: The Epistemological Status of "Structural Prediction"

A claim is a *structural prediction* if it derives from an explicit theoretical model's axioms and parameters, applies to a class of cases defined by the model's primitive concepts, and generates predictions for specific cases that are in principle testable but whose test results do not by themselves confirm or refute the model — only accumulating evidence across many cases constitutes a genuine test. The claim that polycentric emergence is structurally predicted for technologies satisfying Axioms I–V is a structural prediction in this sense: it is not confirmed by the four cases (which were used in constructing the model) and cannot be confirmed by any finite set of cases, but it can be gradually supported or undermined by the accumulation of case studies across a wide range of technologies and periods. Its epistemological status is therefore more like a research programme specification (Lakatos 1978) than like an empirical hypothesis susceptible to single-study confirmation or refutation.

This epistemological characterisation has practical implications for how the series' conclusions should be presented to non-specialist audiences. A claim that "pottery was independently invented five times" is an empirical statement that can be supported or refuted by specific evidence. A claim that "pottery is the kind of technology that is structurally prone to independent re-invention" is a structural prediction whose support is the convergence of axiom-satisfaction evidence, posterior odds, and inter-case consistency documented across this volume. The two claims are related but not equivalent, and conflating them — as popular science treatments of polycentric emergence sometimes do — risks both overstating the certainty of the empirical findings and misrepresenting the theoretical status of the framework.

7.8 The Relationship Between Polycentric Emergence and Human Behavioural Modernity

A set of questions that the Network Horizon framework touches without directly addressing concerns the relationship between the polycentric emergence of Neolithic technologies and debates about the emergence and spread of cognitively modern human behaviour. The technologies examined in this series — ceramic vessels, wheeled transport, horse domestication, copper smelting — are all post-Palaeolithic, and their polycentric patterns are in part a consequence of the demographic conditions of the Holocene: post-glacial population expansion,

agricultural subsistence, and the accumulation of the material prerequisites that bring technologies within the Regime I parameter space. The Palaeolithic, with its sparser populations and denser technological traditions (fire-making, flint-knapping, composite tools), presents a different set of cases where the polycentric versus diffusion question has been debated with substantially less multi-source evidentiary rigour.

Note 7.3: Upper Palaeolithic Technology and the Framework's Applicable Range

The earliest ceramic technology (South China, $c.\ 20{,}000$ BP) sits at the interface between the Upper Palaeolithic and the beginning of the ceramic Neolithic, and the ceramics case is the only core case in the series that substantially engages with Upper Palaeolithic demographic conditions. Extending the framework to Palaeolithic technologies — the production of pigment-based rock art, the development of composite projectile weapons, the manufacture of sewn clothing documented by eyed needles — would require dealing with a substantially degraded evidentiary base (no ancient DNA for most contexts, limited INAA reference databases for ochre and other pigment raw materials, glottochronological analysis well beyond the calibration horizon), and with population densities and contact structures that may not fit cleanly into the Axiom II threshold framework calibrated against Holocene forager communities. The framework is therefore not directly applicable to most Palaeolithic technologies without substantial modification, and claiming its application to debates about Upper Palaeolithic behavioural modernity (the Middle-to-Upper Palaeolithic transition, the possible polycentric origin of blade technology) would require extensions that go beyond what this series can provide.

7.9 A Critical Engagement with the Series' Own Methodology

The penultimate substantive section of the series provides an opportunity for the kind of self-critical methodological assessment that is rarely included in the volumes that constitute an extended research programme, but whose inclusion here is consistent with the series' explicit commitment to pre-registration, transparency about detection limits, and acknowledgment of the limits of the framework.

Case Study 7.1: The Three Methodological Decisions Most Open to Criticism

Decision 1: The multiplicative composite score. The choice to compute $\Sigma(\mathcal{T}) = \prod_k \sigma_k$ rather than an additive combination $\sum_k w_k \sigma_k$ reflects the view that the five axioms are conjunctive requirements — all five must be jointly satisfied for polycentric emergence to be structurally predicted — rather than contributing independently to the probability in a graded way. This choice generates the strong compositional property (a low score on any single axiom dominates the composite) that correctly identifies the wheel's Axiom IV as the binding constraint and copper's Axiom II as the binding constraint. The criticism is that the multiplicative rule may be too sensitive to binding axioms: a score of 0.01 on one axiom reduces the composite by two orders of magnitude even if the other four axioms are all at 0.95, which may overstate the inferential asymmetry between a weak axiom and a strong one. An additive-with-interaction model, in which severe violations of any single axiom dominate the composite but moderate weakness is less catastrophic, might be a better approximation to the true structure of the polycentric emergence condition.

Decision 2: The treatment of the linguistic calibration horizon as a hard cutoff. The glottochronological analysis of Volume IV treats $8{,}000$ years as a hard upper limit beyond which linguistic reconstruction is non-informative. This is a conservative choice that errs in the direction of not overclaiming linguistic evidence, but it is not obviously correct:

the calibration horizon is a reflection of the current state of internal calibration data, not a fundamental limit on linguistic inference. As better-dated genetic and archaeological correlates of language family diversification events are identified, the calibration horizon will move, and treating its current position as definitive may cause systematic underestimation of the linguistic stream's contribution to the ceramics case and some other cases.

Decision 3: The restriction to four cases. The series' scope restriction to four technologies reflects a judgment about evidential completeness (Proposition 1.3) that is defensible but contestable. Including textile production, for which the axiomatic pre-assessment of Finding 6.1 predicts $\hat{\Sigma} \approx 0.55$, would expand the calibration corpus and reduce the prediction error identified in Proposition 7.1, at the cost of applying the framework to a case where one or two evidence streams are currently non-discriminating. Whether the additional calibration value of including a fifth, somewhat weaker case outweighs the dilution of evidential standards is a judgment call that the series made conservatively. Future researchers may reasonably make a different call.

7.10 The Network Horizon as a Model for Multi-Disciplinary Integration

Beyond its specific findings about prehistoric technologies, the Network Horizon series represents a methodological argument: that questions which have traditionally been answered within a single disciplinary framework (archaeological stratigraphy, historical linguistics, archaeometry, population genetics) are better answered by the structured integration of evidence across all four frameworks under a formal Bayesian architecture. The argument is not that any individual discipline's methods are inadequate but that the questions addressed by those methods are genuinely multi-dimensional, and that a formal framework for integrating their outputs is superior to the informal, case-by-case weighing that typically occurs in multi-disciplinary synthesis.

Proposition 7.3: The Information Gain from Multi-Source Integration

Let I_k denote the mutual information between evidence stream $\mathbf{E}_k$ and the binary outcome $(\bar{\lambda} < \rho^*)$ for a typical prehistoric technology polycentric emergence question. For the four core cases of the Network Horizon series, the estimated mutual information values are approximately $I_1 \approx 0.8$ bits (chronology), $I_2 \approx 1.1$ bits (provenance), $I_3 \approx 1.3$ bits (genetics), $I_4 \approx 0.6$ bits (linguistics). The total information in the four streams, under conditional independence, is $\sum_k I_k \approx 3.8$ bits. The information available from any single stream is therefore at most 1.3 bits, leaving ≈ 2.5 bits of evidence — roughly an eight-fold increase in evidential weight — inaccessible to single-discipline analyses. The multi-source framework recovers this additional information by combining streams that each carry independent evidence about different aspects of the inter-regional contact density and independent invention probability.

Note 7.4: The Institutional Challenge of Multi-Source Integration

The information-theoretic argument of Proposition 7.3 establishes that multi-source integration is, in principle, superior to single-source analysis for the Diffusion Boundary Problem. In practice, achieving multi-source integration requires overcoming institutional barriers that the information-theoretic framework does not address: the four evidence streams are produced by researchers trained in different disciplines, publishing in different journals, evaluated by different peer review communities, and operating under

different standards of methodological rigour. The synthesis required by the Network Horizon framework presupposes that a single research team can achieve competence across all four disciplines or that effective collaboration across disciplinary boundaries can be sustained over the timescale of a multi-volume research programme. Neither of these conditions is easily satisfied under current academic institutional structures, which reward disciplinary depth over multi-disciplinary breadth and provide limited mechanisms for sustaining collaborative research projects across the five-to-ten-year timescales required for projects of this scope. The institutional challenge is not specific to the Network Horizon framework; it is a general feature of any research programme that requires genuine integration across disciplinary evidence streams. Noting it here is an acknowledgment that the research programme's continuation depends on institutional as well as intellectual conditions.

7.11 Final Accounting: What Remains Unknown

The seven chapters of this volume have formalised the Network Horizon model's axiomatic structure, applied it retrospectively to four core cases, developed the evidentiary framework for the Diffusion Boundary Problem, synthesised the methodological apparatus, derived the model's falsifiable predictions, and acknowledged its limits. The final task is an honest accounting of what this edifice leaves unknown — not as an apology for the framework's limitations but as a precise specification of the research agenda it opens.

Finding 7.1: The Seven Open Questions in Priority Order

Ranked by their importance for the Network Horizon research programme's next phase, the seven most consequential open questions are:

(1) What is the base rate of polycentric emergence for technologies satisfying Axioms I–V? This requires expanding the calibration corpus beyond four cases.

(2) Can the Sub-Saharan African iron smelting case be resolved by ancient DNA? This is the single case most likely to produce a new strong polycentric result with currently achievable aDNA sampling.

(3) What is the Regime I/II transition date for ceramics in the Near East, where the Natufian interaction sphere provides the most tractable diachronic transition to document?

(4) Can the non-functional choice divergence index $D_{\mathcal{NF}}$ (Definition 5.6) be extended to horse management traditions using morphometric and ancient genomic data as proxies for production-sequence choices?

(5) What is the correct prior odds ratio π_D/π_P for the wheel given the Bronze Age Eurasian contact network, and would a blind pre-assessment protocol (Definition 6.3) yield a different prior than the one implicitly used in Volume II?

(6) Does the candidate Axiom VI (social technology receptivity, Definition 2.4) add discriminating power for technologies not captured by the current five-axiom system, specifically writing systems and standardised coinage?

(7) What is the relationship between the Network Horizon model's contact density parameter $\bar{\lambda}$ and the genetic concept of effective migration rate m_e as estimated by ADMIXTOOLS? A formal equivalence or conversion formula between these two quantities would allow direct use of qpAdm estimates to constrain Axiom IV's contact density condition.

These seven questions are not afterthoughts appended to a completed research programme. They are the precisely defined starting points of the programme that the Network Horizon

series has made possible: questions that could not have been rigorously asked before the formal apparatus of this volume existed, and that are now askable with the degree of precision that their answers require.

7.12 Implications for the Philosophy of Archaeological Inference

Beyond its specific empirical findings and methodological apparatus, the Network Horizon series bears on a broader question in the philosophy of historical and archaeological inference: whether disciplines that cannot perform controlled experiments can nonetheless achieve the kind of principled hypothesis discrimination that experimental sciences take for granted. This concluding section situates the series's contribution within that wider debate.

Note 7.9: The Underdetermination Problem in Historical Science

A standard worry in the philosophy of historical science holds that the historical record radically underdetermines the causal processes that produced it: multiple distinct causal histories can, in principle, generate observationally indistinguishable traces, and no amount of additional excavation can fully resolve this underdetermination because the past, unlike a laboratory system, cannot be re-run under controlled conditions to isolate causal variables. This worry has particular force for the diffusion-versus-independent-invention question, since both processes can in principle produce convergent technological outcomes, and the worry might seem to undermine the entire project of the Network Horizon series. The series's response, implicit throughout but worth making explicit here, is that underdetermination by any single evidentiary stream does not entail underdetermination by the conjunction of multiple independent streams: while chronological evidence alone may be compatible with both diffusion and independent invention for a given case, the joint requirement that chronological, genetic, archaeometric, and linguistic evidence all be simultaneously compatible with a single causal history is a substantially more restrictive condition, and it is this multi-stream joint constraint, rather than any single stream's discriminating power, that does the principal epistemic work of the series.

Proposition 7.4: Multi-Source Convergence as a Partial Response to Underdetermination

For a set of k evidentiary streams each individually subject to the underdetermination worry (each compatible with both hypotheses under some non-trivial range of auxiliary assumptions), the joint probability that all k streams are simultaneously compatible with the disfavoured hypothesis under independently varying auxiliary assumptions decreases multiplicatively with k, provided the streams' auxiliary-assumption spaces are themselves independent (an assumption examined and partially qualified in this chapter's discussion of site-level evidence correlation). This does not eliminate underdetermination in the strict philosophical sense — there remains, for any finite k, a logically possible joint specification of auxiliary assumptions across all k streams that reconciles the evidence with the disfavoured hypothesis — but it renders such joint specifications increasingly contrived and increasingly vulnerable to independent empirical check as k grows, which is the most that historical inference of any kind can achieve and is consistent with how multi-source convergence functions as a confirmatory standard in other historical and forensic disciplines.

7.13 The Series in Relation to Falsificationist and Bayesian Philosophy of Science

The framework's methodology draws on both falsificationist and Bayesian traditions in the philosophy of science without committing fully to either, and this hybrid character is worth examining explicitly as the series concludes.

Corollary 7.3: A Bayesian Framework with Falsificationist Discipline

The Network Horizon framework's core inferential engine — the Bayes factor decomposition of Definition 3.5, the posterior updating discussed throughout Volumes I–IV — is explicitly Bayesian, treating evidence as updating prior probabilities rather than as definitively confirming or refuting hypotheses in the Popperian sense. At the same time, the prospective prediction methodology of Chapter 6, with its emphasis on temporal priority, non-triviality, and risk (Proposition 6.5), imports a falsificationist discipline onto the Bayesian engine: predictions are required to be the kind of claim that could, in principle, come out wrong, and the framework's credibility is treated as resting partly on its track record of risky predictions rather than solely on the internal coherence of its posterior probability calculations. This hybrid stance reflects a broader trend in the philosophy of science toward what has been termed Bayesian confirmation theory with falsificationist constraints: probability updating provides the formal machinery for combining evidence, while falsificationist criteria provide the discipline that prevents the formal machinery from being applied in ways that are immune to genuine empirical risk.

7.14 Anticipated Objections Not Yet Addressed

In the interest of intellectual honesty, this section catalogues objections to the framework that the authors consider serious but have not been able to address fully within the scope of this volume, distinguishing them from the open questions already discussed, which concern extensions and refinements rather than potential weaknesses in the existing argument.

Note 7.5: The Problem of Researcher Degrees of Freedom in Axiom Scoring

Notwithstanding the inter-rater reliability checks reported in Chapter 5, a more fundamental version of the researcher-degrees-of-freedom objection remains only partially addressed: the axiom-scoring protocol, however reliable across assessors operating within the same interpretive framework, was itself developed by researchers who also constructed the cases used to validate it, creating a risk that the entire scoring methodology — not merely individual scores within it — has been shaped, consciously or not, by knowledge of the calibration corpus's outcomes. The negative control of Worked Example 1.3 and the robustness analysis of Chapter 2 partially address this concern but do not eliminate it, since both were themselves constructed using the same scoring protocol under scrutiny. A fully adequate response would require the scoring protocol to be applied by researchers with no involvement in its construction to cases selected by a third party with no stake in the outcome, a validation exercise beyond the scope of what a single-author series can accomplish and one that the authors hope the broader research community will undertake.

7.15 Closing Reflection on the Series as a Whole

The five volumes of the Network Horizon series, taken together, advance a specific and bounded claim: that for a defined set of Neolithic and Chalcolithic technologies, satisfying a specified set of structural conditions, the weight of multi-source archaeological and genetic evidence favours independent invention over contact-mediated diffusion as the explanation for observed cross-regional technological convergence. This claim is narrower than the popular conception of "polycentric origins" debates often suggests, and the axiomatic and evidentiary apparatus developed to support it is more constrained, more falsifiable, and more openly qualified by its own limitations than either strict diffusionism or unconstrained polycentrism as informally practised in older archaeological literature.

The series does not claim to have resolved the diffusion-versus-independent-invention question in general, nor does it claim that its four core findings extend automatically to technologies, regions, or periods outside the calibration corpus. What it claims, and what the cumulative argument of these five volumes is intended to establish, is that the question is empirically tractable when approached with sufficiently rigorous and sufficiently diverse evidence, that the apparent intractability of the question in earlier archaeological debate reflected an evidentiary and methodological gap rather than a genuine epistemic limit, and that closing this gap changes the balance of evidence, for at least these four cases, in a specific and defensible direction. Whether this balance shifts again as new evidence streams become available, as the calibration corpus expands, and as the framework's own axioms are tested against independent scrutiny, is precisely the kind of question a properly falsifiable framework should leave open.

A terminological note before closing: readers arriving at this series from the popular "multiple independent discoveries" narrative in science communication (familiar anecdotes about calculus, the telephone, or natural selection) should not read this series's findings as general support for that looser narrative. The popular version typically describes well-connected scientific communities in active, if indirect, communication — conditions close to the opposite of the demographic-isolation requirement central to Axiom IV here. This series's claims are narrower, more conditional, and rest on a substantially more stringent evidentiary bar than that popular framing usually applies.

7.16 A Final Note on the Relationship Between Formal Rigour and Archaeological Interpretation

The extensive formal apparatus developed across this series — axioms, theorems, propositions, Bayes factor decompositions, profile-distance measures — might appear, to readers more accustomed to traditional archaeological prose, as an imposition of mathematical formalism onto a domain whose evidentiary base (sparse, fragmentary, subject to taphonomic loss and excavation bias) does not obviously warrant the apparent precision such formalism suggests. This concern deserves a direct response as the series concludes.

Corollary 7.4: Formalism as Discipline, Not as Claim to False Precision

The formal apparatus developed throughout this series is not intended to claim a level of numerical precision unwarranted by the underlying archaeological evidence; the explicit uncertainty quantification accompanying every reported composite score, sensitivity analysis, and robustness check (Theorem 2.7, Proposition 2.2, Proposition 5.5, among others) is precisely the mechanism by which the formalism avoids overclaiming precision. The function of the formalism is instead to impose interpretive discipline: by re-

quiring that intuitive judgements about plausibility, redundancy, threshold accessibility, and contact-network density be made explicit, quantified (even approximately), and combined according to a stated and defensible logical structure, the framework makes those judgements available for scrutiny, replication, and revision in a way that purely narrative archaeological argument, however expert, typically does not. A formal framework that turns out, on scrutiny, to rest on the same underlying expert judgements that an experienced archaeologist would bring to an unformalised assessment has still made a contribution, provided it has made those judgements visible, checkable, and revisable rather than leaving them implicit in the prose of an interpretive narrative. This is the standard against which the present series's formal apparatus should be judged, and it is a standard the series aims to meet not by claiming false numerical precision but by making its inferential structure as transparent as the available evidence permits.

Note 7.8: An Invitation to Replication and Extension

In keeping with the transparency commitment articulated in Corollary 7.4 and the audit-trail recommendations of Chapter 5, the authors close this volume with an explicit invitation to other researchers: to apply the axiomatic scoring protocol independently to the four core cases and report any divergence from the scores reported here; to extend the calibration corpus to technologies, regions, and periods outside the present scope, following the individuation and evidential-completeness criteria of Chapter 1; and to subject the framework's combination rules, robustness properties, and falsifiability criteria to the kind of independent scrutiny that no single research programme, however extensive, can fully provide for itself. The Network Horizon framework is offered not as a finished theory but as a falsifiable, extensible apparatus whose value will ultimately be determined by how it performs under exactly this kind of independent test.

7.17 Toward a General Theory of Technological Convergence: Speculative Extensions Beyond Archaeology

The Network Horizon framework, as developed across this volume, is constructed specifically for the deep-time archaeological problem of distinguishing diffusion from independent invention in prehistoric technologies. Its formal apparatus, however — a threshold condition separating regimes of structural convergence and divergence, derived from a small set of axiomatic preconditions and tested against multi-source evidence — has a level of generality that invites speculation about extension to domains well outside archaeology, where structurally analogous questions about convergent versus contact-mediated similarity arise. This closing section sketches such extensions explicitly as speculation, clearly demarcated from the volume's empirical claims, in the spirit of the open-questions programme of Chapter 7.

Proposition 7.6: Structural Analogy to Convergent Evolution in Biology

The distinction between polycentric technological emergence and diffusion has a long-recognised structural analogue in evolutionary biology's distinction between convergent evolution (independent lineages arriving at similar functional solutions, such as the camera eye in vertebrates and cephalopods) and homologous inheritance (shared structure due to common descent). The analogy is not merely metaphorical: both domains require a method for distinguishing functional similarity that arose independently from functional similarity that arose through a shared causal pathway, and both domains have developed formal apparatus for this purpose — molecular phylogenetics and homology test-

ing in biology, the axiomatic and multi-source evidentiary framework of this volume in archaeology. Axiom V's functional-convergence-without-structural-homology criterion is, in fact, directly borrowed in spirit from the biological convergent-evolution literature's standard diagnostic (independently evolved structures solving the same functional problem through different developmental and structural pathways), and a more systematic comparison of the statistical and inferential machinery developed in each field could plausibly yield methodological cross-fertilisation in both directions, a possibility noted but not pursued in depth in this volume.

Note 7.10: A Cautionary Note on the Limits of the Biological Analogy

The analogy sketched in Proposition 7.6 should not be pressed too far. Biological convergent evolution operates over timescales and through mechanisms (genetic mutation and natural selection acting on heritable variation) that differ fundamentally from the cognitive and cultural mechanisms underlying technological invention, and the formal apparatus developed for molecular phylogenetics — built on assumptions about mutation rates, selection coefficients, and population genetics that have no direct technological analogue — cannot simply be imported wholesale into the archaeological domain. The Network Horizon framework's own axiomatic structure was developed independently, grounded in the specific evidentiary affordances of archaeology, genetics, and linguistics rather than adapted from biological phylogenetic methods, and the analogy noted in Proposition 7.6 is offered as a source of conceptual orientation and a possible direction for future methodological dialogue between fields, not as a claim that biological methods could be directly transplanted to resolve any of the open questions catalogued elsewhere in this chapter.

Proposition 7.7: A Speculative Application to the History of Mathematical and Scientific Discovery

A further, more speculative extension concerns the history of mathematical and scientific ideas, where the question of independent discovery versus diffusion arises with some regularity — the Newton–Leibniz calculus priority dispute being the most famous instance, but comparable questions arise for the independent development of non-Euclidean geometry by Gauss, Bolyai, and Lobachevsky, and for numerous other cases of apparently simultaneous discovery catalogued in the sociology-of-science literature on "multiples" (Merton 1961). These cases differ from the archaeological cases of this volume in evidentiary character — written correspondence, publication dates, and documented institutional contact networks provide a far richer and more direct evidentiary basis than the indirect material, genetic, and linguistic proxies available for prehistoric technologies — but the underlying structural question (does the documented network of scholarly contact between region or community R_i and R_j exceed the threshold required to make diffusion the more parsimonious explanation for a shared discovery, given the discovery's accessibility from existing mathematical knowledge in each community) has the same logical form as the Diffusion Boundary Problem of Chapter 4. Constructing a Network Horizon-style framework for the history of mathematics, with ρ^* estimated from documented correspondence networks rather than archaeological contact-density proxies, is identified here as a speculative but methodologically tractable extension that future work, potentially outside the present author's own research programme, might pursue.

Corollary 7.5: These Extensions Are Offered as Invitations, Not Commitments

The speculative extensions sketched in this closing section are offered in the same spirit as the prospective technological cases of Chapter 6 and the open questions catalogued throughout Chapter 7: as an invitation to researchers in adjacent fields to test whether the formal structure developed here for a specific archaeological problem has genuine application beyond its original domain, rather than as a commitment on the part of this series to pursue those applications itself. The Network Horizon series, across its five volumes, has remained focused on the four core prehistoric technologies and the methodological apparatus required to assess them; extending that apparatus to the history of science, to biological convergent evolution in a more formal comparative sense, or to other domains where independent-origin questions arise, would constitute a substantially different research programme, and the authors make no claim to have begun it here beyond the brief speculative gestures of this section.

Note 7.11: Closing Reflection on the Series as a Whole

The five volumes of the Network Horizon series began with four separate empirical investigations into specific prehistoric technologies and arrived, in this concluding volume, at a unified axiomatic and evidentiary framework that explains why those four investigations reached structurally similar conclusions through what initially appeared to be four independent lines of case-specific argument. That the unification was possible at all — that a single set of five axioms, a single diffusion-boundary formalism, and a single multi-source evidentiary protocol could be retrofitted across cases as materially and chronologically diverse as ceramic technology, copper metallurgy, horse domestication, and wheeled-vehicle technology — is itself a finding worth dwelling on, independent of whether the specific polycentric conclusions reached for each case withstand the scrutiny the falsifiability programme of Chapter 6 invites. Whether the unification reflects a genuine underlying structural regularity in how technologies emerge in demographically isolated populations, or whether it reflects, at least in part, the kind of retrospective theoretical fitting that Definition 1.7's account of research infrastructure artifacts warns against, is a question this series cannot answer from within its own four-case evidentiary basis. It is a question for the research programme this volume opens, and for the researchers, inside and outside the present author's own work, who take up its falsifiable predictions and test them against evidence not yet gathered.

7.18 A Final Methodological Postscript: On Writing a Synthesis Volume After Four Case-Specific Volumes

This concluding chapter has, throughout, addressed substantive open questions and limitations of the Network Horizon framework. A final, more reflective point concerns the methodological experience of constructing this synthesis volume itself, after the four case-specific volumes it draws together, and what that experience suggests about the relationship between case-specific empirical research and subsequent theoretical unification more generally.

Note 7.12: The Risk of Retrospective Coherence

Constructing a unified axiomatic framework after the fact, from four case studies that were originally developed with their own case-specific formal apparatus rather than designed from the outset to instantiate a common theory, carries an inherent risk: the uni-

fication process itself involves choices — which aspects of each case-specific analysis to treat as instances of a general axiom, which to treat as case-specific detail outside the general framework's scope — that could, even with the best intentions, be made in ways that exaggerate the underlying coherence across cases relative to what a researcher starting from the four cases with no prior expectation of unification would find. This risk is structurally related to, but distinct from, the confirmation-bias concern addressed in Definition 1.7: it is not a risk that the case-selection or evidence-gathering was biased toward favourable outcomes, but a risk that the theoretical synthesis itself imposes more apparent order on genuinely independent case-specific findings than those findings, examined without the unifying ambition, would support.

Proposition 7.8: Partial Mitigation Through the Independence of the Original Case-Specific Formal Apparatus

The retrospective-coherence risk of Note 7.12 is partially, though not completely, mitigated by the fact that the four case-specific formal apparatus systems synthesised in this volume — the loanword parsimony test of Volume IV, the diffusion-distance framework of Volume III, the cost-substitution model of Volume II, and the comparative archaeometric framework of Volume I — were each developed independently, prior to and without reference to the unification project that became this volume, in response to the specific evidentiary challenges of their respective cases. The fact that these four independently developed apparatus systems could be retrospectively organised under a common five-axiom structure with a unified diffusion-boundary formalism is therefore not guaranteed by the unification process itself in the way it would be if the four volumes had been planned from the outset as instantiations of a pre-specified general theory; some genuine cross-case structural commonality must have been present in the original four analyses for the unification to have been possible at all, even granting that the specific axiomatic form in which that commonality is here expressed reflects choices made during the synthesis process itself.

Corollary 7.6: The Appropriate Epistemic Status of This Volume's Unification

Taken together, Note 7.12 and Proposition 7.8 suggest the appropriate epistemic status for readers to assign to this volume's central theoretical contribution: the five-axiom framework and diffusion-boundary formalism should be treated as a genuine, falsifiable theoretical proposal — supported by, but not simply read off from, the four case-specific analyses it synthesises — rather than as a neutral redescription of findings that were already fully present and uncontroversial in Volumes I–IV. Readers who find the unification compelling should recognise that they are endorsing a substantive theoretical claim, with the falsifiable content articulated throughout Chapter 6, not merely agreeing that four independent case studies happened to reach broadly similar conclusions. This distinction matters for how the volume's claims should be tested and, if warranted, revised by future research.

Note 7.13: Acknowledgement of the Provisional Nature of This Synthesis

In closing, it is worth restating plainly what the preceding four sections of this chapter have argued from several different angles: this volume offers a provisional synthesis, built from a small corpus of cases, subject to the methodological limitations catalogued throughout Chapter 7, and explicitly designed to generate falsifiable predictions rather than to declare the diffusion-versus-polycentricity question settled. The five axioms,

the diffusion-boundary formalism, and the multi-source evidentiary protocol developed across this volume constitute the most rigorous and explicit statement the present author is currently able to offer of why ceramic technology, copper metallurgy, horse domestication, and wheeled-vehicle technology each independently arose in multiple regions during the relevant prehistoric periods. Whether this statement survives the scrutiny of an expanded calibration corpus, independent replication of its scoring protocols, and the discovery-contingent tests sketched in Chapter 6, is a matter the present volume leaves, deliberately and as a matter of principle, to future research.

Appendix A

Consolidated Axiomatic Profiles of the Four Core Cases

Chapter 3 assesses the four core cases against the five axioms of Chapter 2 axiom by axiom, case by case, with each assessment embedded in several paragraphs of supporting argument. This appendix consolidates those assessments into a single reference table, reproducing only the qualitative satisfaction ratings and composite scores already established in Chapter 3 and cited elsewhere in this volume — it introduces no new evidence or scores. Its purpose is consultative: a reader who needs to compare the four cases' profiles at a glance, after having read the supporting argument in Chapter 3, should not have to re-locate eleven separate paragraphs to do so.

A.1 Qualitative Satisfaction Ratings by Axiom

Axiom	Ceramics	Horse Dom.	Copper	Wheel
I. Material Redundancy	strong	moderate–strong	strong	moderate–strong
II. Low Min. Viable Threshold	strong	moderate–strong	moderate	moderate–strong
III. Recurrent Functional Pressure	strong	strong	strong	moderate–strong
IV. Demographic Isolation	strong*	strong	strong	moderate
V. Functional Convergence w/o Homology	strong	moderate–strong	strong	moderate
Overall profile (Ch. 3)	uniformly strong	moderately strong	moderately strong	weak (boundary)

Table A.1: Axiom-by-axiom qualitative satisfaction ratings for the four core cases, as established in Findings 3.1–3.4. *Subject to the residual ancient-DNA coverage caveat noted in Finding 3.1's Axiom IV discussion.

A.2 Composite Scores Cited in This Volume

Case	$\Sigma(\mathcal{T})$	Source in this volume
Ceramics	0.69	Section 3.5, Comparative Synthesis
Horse domestication	0.89	Section 2.10, Proposition 2.3; Section 4.9, Worked Example 4.11
Copper metallurgy	0.76–0.81	Section 2.10, Proposition 2.3 (sensitivity range; no single point estimate is reported elsewhere in this volume)
The wheel	0.15	Section 3.6, Note 3.2

Table A.2: Composite scores as cited at their respective points in this volume. No table consolidating these four figures appears elsewhere in the text.

Note A.1: What This Table Does Not Claim

Three things are worth stating explicitly about Table A.2 so that its convenience is not mistaken for a precision the underlying numbers do not have. First, the copper figure is a range, not a point estimate, because the case-specific chapters never report a single $\Sigma(\text{copper})$ value; the range given is the sensitivity-analysis interval of Section 2.10, reported here for completeness rather than collapsed to a misleadingly precise midpoint. Second, as Note 2.1b states at the point each case is first scored and as this volume repeats wherever a composite score is used substantively, every σ_k and every $\Sigma(\mathcal{T})$ in this table is an expert-elicited ordinal judgement on a $[0, 1]$ scale, not a measured quantity; the table's numerical alignment should not be read as implying that the underlying axiom-level scores $(\sigma_1, \ldots, \sigma_5)$ for each case are individually available in a form precise enough to be tabulated — only the qualitative ratings of Table A.1 and the aggregate Σ figures of Table A.2 are reported anywhere in this volume, and this appendix reproduces exactly that and no more. Third, the four figures are not on a forced common scale across genuinely independent evidentiary bases: each $\Sigma(\mathcal{T})$ is internal to its own case's evidentiary profile, and the cross-case comparison they permit is ordinal (which case is more or less evidentially secure) rather than cardinal (precisely how many times more secure).

Appendix B

A Candidate Sixth Axiom: Institutional Prerequisite

Section 2.6 notes, in a single paragraph, that the five-axiom system developed in Chapter 2 abstracts away from the internal social and institutional dynamics of invention and adoption, and that a fuller theory would require integrating the Network Horizon model's material and demographic axes with an account of institutional evolution — a project flagged there as beyond the volume's scope. This appendix takes that one-paragraph flag and develops it far enough to show concretely why the candidate axiom was not adopted, rather than leaving the reader to take the exclusion on faith.

B.1 Motivation: A Gap the Five Axioms Do Not Cover

Axioms I–V jointly specify material, cognitive, motivational, and demographic conditions for polycentric emergence. None of them addresses a fifth consideration that matters for some technologies even when Axioms I–V are satisfied: whether producing $\mathcal{T}$ requires not merely individual or small-group competence (Axiom II's concern) but a standing social institution — a role, an office, a controlled apprenticeship, or an exclusive lineage of practice — that must already exist in a region before an individual invention event of the kind Definition 2.1 describes can occur at all.

Definition B.1: Institutional Prerequisite

A technology $\mathcal{T}$ has *institutional prerequisite* if its minimum viable threshold $\theta(\mathcal{T})$ is achievable by a single competent individual or small group (Axiom II is satisfied), but its *sustained reproduction* beyond a single invention event requires a standing social institution $\mathcal{I}(R)$ — a specialist role, a restricted-access apprenticeship structure, or an exclusive lineage of transmission — that is not guaranteed to exist merely because the material, cognitive, and demographic conditions of Axioms I, III, and IV are met. Formally, let $\pi(\mathcal{T}, R)$ denote the probability that a single successful invention event $\iota(R, \mathcal{T}, t)$ is followed by sustained local reproduction of $\mathcal{T}$ within R rather than the technique being lost within one or two generations. Institutional prerequisite holds when $\pi(\mathcal{T}, R)$ depends materially on $\mathcal{I}(R)$'s presence, independent of $\theta(\mathcal{T})$.

None of the four core cases of this series exhibits strong institutional prerequisite in this sense: ceramic production, horse handling, native-copper cold-working, and basic cart construction are all, once invented, reproducible by ordinary intergenerational observation and practice without a specialist institutional structure standing between invention and persistence. This is precisely why Section 2.6 could note the gap and set it aside: the four core cases

do not require it. The candidate axiom becomes relevant for technologies, including some of Chapter 6's prospective cases, where invention and persistence may come apart.

B.2 Why Iron Smelting Suggests the Need for the Distinction

Chapter 6 already identifies iron smelting as a Regime II-leaning case on Axiom II grounds: the $1{,}200^{\circ}$C-plus sustained-temperature requirement with carburation control sits above the forager threshold θ_{forager}. Definition B.1 suggests a sharper version of that diagnosis. Even granting, for the sake of argument, a region where a sufficiently skilled individual smith independently solves the technical problem of bloomery iron production — satisfying $\theta(\mathcal{T}) \leq \theta_{\text{forager}}$ for that individual, contrary to the Chapter 6 assessment — the historically documented pattern of early iron-working traditions is one of restricted, often hereditary or caste-linked, smithing roles (Childs and Killick 1993; Schmidt and Avery 1978 for ethnographically documented sub-Saharan African cases). This suggests that even where Axiom II might be marginally satisfiable by an exceptional individual, $\pi(\mathcal{T}, R)$ for iron smelting may depend on the prior or concurrent existence of an institutional structure capable of protecting, transmitting, and socially valorising specialist knowledge across generations — a condition orthogonal to Axiom II's individual-competence threshold.

Proposition B.1: Institutional Prerequisite as a Compounding, Not Replacing, Constraint

For a technology $\mathcal{T}$ exhibiting institutional prerequisite in the sense of Definition B.1, the effective probability of polycentric emergence is not the five-axiom composite $\Sigma(\mathcal{T})$ alone but a further-discounted quantity $\Sigma(\mathcal{T}) \cdot \pi(\mathcal{T}, R)$, where $\pi(\mathcal{T}, R) < 1$ whenever $\mathcal{J}(R)$ cannot be assumed present independently of the invention event itself. This is a compounding constraint, not a replacement for Axiom II: a technology could in principle satisfy Axiom II (an individual can solve the technical problem) while still having low $\pi(\mathcal{T}, R)$ (the solution does not persist locally absent a supporting institution), a combination Axiom II alone cannot detect because Axiom II is defined purely at the level of individual or small-group technical competence.

B.3 Why This Is Offered as an Appendix, Not a Sixth Axiom in Chapter 2

Three considerations, taken together, explain why Definition B.1 and Proposition B.1 are developed here rather than promoted into the main axiomatic system of Chapter 2.

Note B.1: Three Reasons Against Promotion

First, $\pi(\mathcal{T}, R)$ as defined in Definition B.1 has not been operationalised against any of the four core cases or the prospective cases of Chapter 6 with the kind of evidentiary protocol Chapter 5 establishes for Axioms I–V; it is, at this stage, a conceptual distinction rather than a scored variable, and adding an unscored sixth axiom to the formal system of Chapter 2 would create an asymmetry between a fully operationalised five-axiom core and a sixth axiom that exists only in definitional form. Second, the institutional-prerequisite question may not be a single binary or graded condition comparable in kind to Axioms I–V, but rather a qualitatively different sort of variable — closer to a moderating condition on Axiom II's reproduction dynamics than to a freestanding necessary condition of the same logical type as material redundancy or demographic isolation; folding it into the same multiplicative composite $\Sigma(\mathcal{T}) = \prod_k \sigma_k$ alongside Axioms I–V, as Proposition

B.1's notation might superficially suggest, risks treating two logically distinct kinds of condition (preconditions for a single invention event versus preconditions for that event's persistence) as though they were the same kind of thing. Third, and most simply, none of the four core cases needs it: Section 2.6's one-paragraph treatment was proportionate to a consideration that does not change any of this volume's four classifications, and a full sixth-axiom apparatus would add formal weight disproportionate to its evidentiary payoff for the cases this volume actually examines.

Corollary B.1: The Appropriate Status of This Appendix

Definition B.1 and Proposition B.1 are offered as a developed conceptual tool for future application to technologies, such as iron smelting and other specialist-craft cases, for which institutional prerequisite plausibly matters, not as an addition to the Chapter 2 axiomatic system governing this volume's own four-case and prospective-case analyses. Should a future case-specific study operationalise $\pi(\mathcal{T}, R)$ against a real evidentiary protocol — for instance, using ethnoarchaeological and historical data on craft-specialist social organisation, in the spirit of the ethnoarchaeological integration discussed in Chapter 5 — and find that it materially changes classification outcomes Axioms I–V alone would not predict, promoting institutional prerequisite to a sixth formal axiom in a subsequent volume would be warranted. Until then, it remains, deliberately, an appendix rather than a chapter.

References

Anthony, D. W. (2007). *The Horse, the Wheel, and Language: How Bronze-Age Riders from the Eurasian Steppes Shaped the Modern World*. Princeton University Press.

Anthony, D. W. (2019). Population genomics and archaeological steppe history. *Journal of Archaeological Research*, 27(3), 287–325.

Binford, L. R. (1962). Archaeology as anthropology. *American Antiquity*, 28(2), 217–225.

Boyd, R., & Richerson, P. J. (1985). *Culture and the Evolutionary Process*. University of Chicago Press.

Burmeister, S. (2004). Der Wagen im Neolithikum und der Bronzezeit. *Mitteilungen der Berliner Gesellschaft für Anthropologie, Ethnologie und Urgeschichte*, 25, 13–40.

Chase-Dunn, C., & Hall, T. D. (1991). *Core/Periphery Relations in Precapitalist Worlds*. Westview Press.

Childs, S. T., & Killick, D. (1993). Indigenous African metallurgy: Nature and culture. *Annual Review of Anthropology*, 22, 317–337.

Clarke, D. L. (1968). *Analytical Archaeology*. Methuen.

Englund, R. K. (1998). Texts from the Late Uruk period. In *Mesopotamien: Späturuk-Zeit und Frühdynastische Zeit*, 15–233. Universitätsverlag Freiburg.

Feldman, M. W. (with Odling-Smee, F. J., & Laland, K. N.) (2003). *Niche Construction: The Neglected Process in Evolution*. Princeton University Press.

Gosselain, O. P. (1992). Technology and style: Potters and pottery among Bafia of Cameroon. *Man*, 27(3), 559–586.

Haak, W., et al. (2015). Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature*, 522, 207–211.

Haarmann, H. (2016). *Ancient Knowledge, Ancient Know-How, Ancient Reasoning*. Cambridge Scholars Publishing.

Henrich, J. (2016). *The Secret of Our Success: How Culture Is Driving Human Evolution, Domesticating Our Species, and Making Us Smarter*. Princeton University Press.

Henrich, J., & McElreath, R. (2003). The evolution of cultural evolution. *Evolutionary Anthropology*, 12(3), 123–135.

Hodder, I. (1982). *Symbols in Action: Ethnoarchaeological Studies of Material Culture*. Cambridge University Press.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90(430), 773–795.

Kloekhorst, A. (2008). *Etymological Dictionary of the Hittite Inherited Lexicon*. Brill.

Larson, G., et al. (2014). Current perspectives and the future of domestication studies. *Proceedings of the National Academy of Sciences*, 111(17), 6139–6146.

Lemonnier, P. (1992). *Elements for an Anthropology of Technology*. University of Michigan Museum of Anthropology.

Livingstone Smith, A. (2001). Bonfire II: The return of pottery firing temperatures. *Journal of Archaeological Science*, 28(9), 991–1003.

Lycett, S. J. (2009). Understanding ancient hominin dispersals using artefactual data: A phylogeographic analysis of Acheulean handaxes. *PLOS ONE*, 4(10), e7404.

Mathieson, I., et al. (2015). Genome-wide patterns of selection in 230 ancient Eurasians. *Nature*, 528, 499–503.

Mauss, M. (1935). Les techniques du corps. *Journal de Psychologie*, 32(3–4), 271–293.

Maynard Smith, J. (1982). *Evolution and the Theory of Games*. Cambridge University Press.

Merton, R. K. (1961). Singletons and multiples in scientific discovery. *Proceedings of the American Philosophical Society*, 105(5), 470–486.

O'Brien, M. J., & Lyman, R. L. (2000). *Applying Evolutionary Archaeology: A Systematic Approach*. Kluwer Academic/Plenum.

Outram, A. K., et al. (2009). The earliest horse harnessing and milking. *Science*, 323(5919), 1332–1335.

Renfrew, C. (1973). *Before Civilization: The Radiocarbon Revolution and Prehistoric Europe*. Jonathan Cape.

Richerson, P. J., & Boyd, R. (2005). *Not by Genes Alone: How Culture Transformed Human Evolution*. University of Chicago Press.

Roux, V. (2019). *Ceramics and Society: A Technological Approach to Archaeological Assemblages*. Springer.

Roux, V., & Corbetta, D. (1990). *The Potter's Wheel: Craft Specialization and Technical Competence*. Oxford & IBH Publishing.

Schmidt, P. R., & Avery, D. H. (1978). Complex iron smelting and prehistoric culture in Tanzania. *Science*, 201(4361), 1085–1089.

Shanks, M., & Tilley, C. (1987). *Re-Constructing Archaeology: Theory and Practice*. Cambridge University Press.

Skibo, J. M., & Schiffer, M. B. (2008). *People and Things: A Behavioral Approach to Material Culture*. Springer.

Skyrms, B. (1996). *Evolution of the Social Contract*. Cambridge University Press.

Thomason, S. G., & Kaufman, T. (1988). *Language Contact, Creolization, and Genetic Linguistics*. University of California Press.

Wallerstein, I. (1974). *The Modern World-System I: Capitalist Agriculture and the Origins of the European World-Economy in the Sixteenth Century*. Academic Press.

Zeder, M. A. (2011). The origins of agriculture in the Near East. *Current Anthropology*, 52(S4), S221–S235.

Zeder, M. A. (2012). The domestication of animals. *Journal of Anthropological Research*, 68(2), 161–190.